

日本国特許庁
JAPAN PATENT OFFICE

10.822.725

07.20.04

別紙添付の書類に記載されている事項は下記の出願書類に記載されて
いる事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed
with this Office.

出願年月日 2004年 2月 5日
Date of Application:

出願番号 特願2004-029028
Application Number:
[JP2004-029028]
ST. 10/C]:

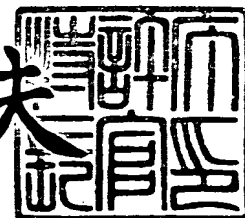
願人 株式会社日立製作所
Applicant(s):

CERTIFIED COPY OF
PRIORITY DOCUMENT

2004年 4月21日

特許庁長官
Commissioner,
Japan Patent Office

今井康夫



【書類名】 特許願
【整理番号】 340301796
【あて先】 特許庁長官殿
【国際特許分類】 G06F 03/06
【発明者】
 【住所又は居所】 神奈川県小田原市中里 3 2 2 番 2 号 株式会社日立製作所 R A I
 D システム事業部内
 【氏名】 星野 幸子
【発明者】
 【住所又は居所】 神奈川県小田原市中里 3 2 2 番 2 号 株式会社日立製作所 R A I
 D システム事業部内
 【氏名】 坂口 孝
【発明者】
 【住所又は居所】 神奈川県小田原市中里 3 2 2 番 2 号 株式会社日立製作所 R A I
 D システム事業部内
 【氏名】 長副 康之
【発明者】
 【住所又は居所】 神奈川県小田原市中里 3 2 2 番 2 号 株式会社日立製作所 R A I
 D システム事業部内
 【氏名】 杉野 昇史
【特許出願人】
 【識別番号】 000005108
 【氏名又は名称】 株式会社日立製作所
【代理人】
 【識別番号】 100095371
 【弁理士】
 【氏名又は名称】 上村 輝之
【選任した代理人】
 【識別番号】 100089277
 【弁理士】
 【氏名又は名称】 宮川 長夫
【選任した代理人】
 【識別番号】 100104891
 【弁理士】
 【氏名又は名称】 中村 猛
【手数料の表示】
 【予納台帳番号】 043557
 【納付金額】 21,000円
【提出物件の目録】
 【物件名】 特許請求の範囲 1
 【物件名】 明細書 1
 【物件名】 図面 1
 【物件名】 要約書 1
 【包括委任状番号】 0110323

【書類名】 特許請求の範囲**【請求項 1】**

上位装置とのデータ授受をそれぞれ制御する複数のチャネルアダプタと、
論理的な記憶領域をそれぞれ提供する複数の記憶デバイス群と、
前記各記憶デバイス群とのデータ授受をそれぞれ制御する複数のディスクアダプタと、
前記各チャネルアダプタ及び前記各ディスクアダプタによりそれぞれ使用されるキャッシュメモリと、
前記キャッシュメモリを論理的に分割して構成される複数のキャッシュ分割領域と、
前記各キャッシュ分割領域を管理するための管理情報を記憶する制御メモリと、
を備え、
前記管理情報は、前記各キャッシュ分割領域毎にそれぞれ設けられる分割管理情報と、
前記各キャッシュ分割領域の全体に適用される共通管理情報とから構成されるストレージサブシステム。

【請求項 2】

前記各分割管理情報及び前記共通管理情報は、キャッシュ管理単位の属性に基づいて設定される請求項 1 に記載のストレージサブシステム。

【請求項 3】

前記管理情報は、複数種類のサブ管理情報から構成されており、
キャッシュ管理単位の属性に基づいて、前記各サブ管理情報の一部を前記各キャッシュ分割領域毎にそれぞれ分割して前記各分割管理情報を構成し、前記各サブ管理情報の残りを前記共通管理情報として使用する請求項 1 に記載のストレージサブシステム。

【請求項 4】

前記管理情報は、複数種類のキューと、これら各キューにそれぞれ関連付けられたカウンタとを含んで構成されており、
キャッシュ管理単位の属性に基づいて、前記各キュー及び前記各カウンタの一部を前記各キャッシュ分割領域毎にそれぞれ設けることにより前記各分割管理情報を構成し、前記各キューの残り及び前記各カウンタの残りを前記共通管理情報として使用する請求項 1 に記載のストレージサブシステム。

【請求項 5】

互いに関連付けられている前記キューと前記カウンタのうち、一方は前記分割管理情報を構成し、他方は前記共通管理情報として使用される請求項 4 に記載のストレージサブシステム。

【請求項 6】

前記各キューにはそれぞれキュー管理テーブルが関連付けられており、前記分割管理情報を構成するキューに関連付けられるキュー管理テーブルは、前記各キャッシュ分割領域毎にそれぞれ設けられる請求項 4 に記載のストレージサブシステム。

【請求項 7】

前記管理情報は、未使用状態のキャッシュ管理単位が接続されるフリーキュー及びこのフリーキューに関連付けられるフリーキューカウンタと、前記記憶デバイス群に反映前のダーティ状態のデータを格納するキャッシュ管理単位が接続されるダーティキュー及びこのダーティキューに関連付けられるダーティキューカウンタと、前記記憶デバイス群に反映済のクリーン状態のデータを格納するキャッシュ管理単位が接続されるクリーンキュー及びこのクリーンキューに関連付けられるクリーンキューカウンタと、使用中のキャッシュ管理単位の総量をカウントする使用中カウンタとを含んで構成され、

前記フリーキューカウンタと、前記クリーンキューと、前記クリーンキューカウンタ及び前記使用中カウンタは、前記各キャッシュ分割領域毎にそれぞれ設けられて、前記分割管理情報をそれぞれ構成し、

前記フリーキューと、前記ダーティキュー及び前記ダーティキューカウンタは、前記共通管理情報として使用される請求項 1 に記載のストレージサブシステム。

【請求項 8】

前記各チャネルアダプタ毎に前記各キャッシュ分割領域をそれぞれ設定可能な請求項 1 に記載のストレージサブシステム。

【請求項 9】

前記各キャッシュ分割領域のうち 1 つのキャッシュ分割領域を共用領域として設定し、この共用領域に属する資源を割り当てることにより、新たなキャッシュ分割領域を設定する請求項 1 に記載のストレージサブシステム。

【請求項 1 0】

上位装置とのデータ授受をそれぞれ制御する複数の上位インターフェース制御部と、論理的な記憶領域をそれぞれ提供する複数の記憶デバイス群と、

前記各記憶デバイス群とのデータ授受をそれぞれ制御する複数の下位インターフェース制御部と、

前記各上位インターフェース制御部及び前記各下位インターフェース制御部によりそれぞれ使用されるメモリ部と、を備えたストレージサブシステムを制御するための方法であって、

前記メモリ部が提供するキャッシュ領域を複数のキャッシュ分割領域に分割し、

前記メモリ部でデータを管理するためのキャッシュ管理単位の属性に応じて、管理情報を前記各キャッシュ分割領域毎にそれぞれ分割し、

前記各管理情報に基づいて、前記各キャッシュ分割領域毎にそれぞれデータを管理させるストレージサブシステムの制御方法。

【書類名】明細書**【発明の名称】** ストレージサブシステム及びストレージサブシステムの制御方法**【技術分野】****【0 0 0 1】**

本発明は、ストレージサブシステム及びストレージサブシステムの制御方法に関する。

【背景技術】**【0 0 0 2】**

ストレージサブシステムは、ホストコンピュータに対して大容量かつ高性能のストレージサービスを提供可能である。ストレージサブシステムでは、例えば、多数のディスクドライブをアレイ状に配設し、RAID (Redundant Array of Independent Inexpensive Disks) に基づく記憶領域を構築している。各ディスクドライブが有する物理的な記憶領域上には、論理的な記憶領域である論理ボリュームが形成されている。論理ボリュームには、LUN (Logical Unit Number) が予め対応付けられている。ホストコンピュータは、LUNやブロックアドレス等を特定することにより、ストレージサブシステムに対して所定形式の書込みコマンド又は読出しコマンドを発行する。これにより、ホストコンピュータは、ストレージサブシステムに対して所望のデータの読み書きを行うことができる。

【0 0 0 3】

ストレージサブシステムには複数のホストコンピュータを接続可能である。あるホストコンピュータまたはアプリケーションプログラムが管理するデータ群を、他のホストコンピュータまたはアプリケーションプログラムから読み書き可能である場合、不都合を生じることがある。そこで、例えば、ゾーニングやLUNマスキング等のようなアクセス制御技術が用いられる。ゾーニングとは、ストレージサブシステムに1つまたは複数のゾーンを設定し、ゾーンに属する特定の通信ポートやWWN (World Wide Name) のみにデータ転送を許可する技術である。LUNマスキングとは、特定のホストコンピュータに対して特定のLUNへのアクセスを許可する技術である。

【0 0 0 4】

ところで、ホストコンピュータとディスクドライブとの間のデータ授受は、キャッシュメモリを介して行われる。ホストコンピュータからライトコマンドが出力されると、例えば、データはいったんキャッシュメモリに保存された後で、ディスクドライブに書き込まれる。また、ホストコンピュータからリードコマンドが出力されると、例えば、ディスクドライブから読み出されたデータは、キャッシュメモリを介してホストコンピュータに提供される。そこで、ディスクドライブへのキャッシュメモリ割当量を適切に配分させる技術も知られている(特許文献1)。

【特許文献1】 特開平4-264940号公報**【発明の開示】****【発明が解決しようとする課題】****【0 0 0 5】**

従来のゾーニングやLUNマスキングにより、ホストコンピュータからストレージサブシステムへのアクセス制限を設定することができる。しかし、従来技術では、論理ボリュームへの単純なアクセス制限が可能となるだけで、ストレージサブシステムが有する資源の分割管理を実現するものではないから、使い勝手が悪い。複数ユーザによってストレージサブシステムが共同で使用されている場合に、例えば、あるユーザが多量の入出力要求(I/O要求)を出すと、キャッシュメモリの多くの領域がこの要求の処理のために使用される。従って、他のユーザからの入出力処理に十分なキャッシュメモリを使用することができなくなり、サブシステム全体としてのサービス性能が低下する。つまり、あるユーザへのサービスのために、他のユーザへのサービスに影響が及ぶ。

【0 0 0 6】

なお、上記特許文献では、各記憶装置にそれぞれキャッシュメモリを割り当てているが、これはキャッシュメモリに記憶されたデータを効率的に記憶装置に書き込むためのもの

であり、複数ユーザに提供するために資源（キャッシュメモリ）の分割管理を行うものではない。

【0007】

そこで、本発明の一つの目的は、複数のユーザにそれぞれキャッシュ領域を割り当てて共同で使用させることができ、あるユーザへ提供するサービスのために他のユーザへ影響が及ぶのを低減させることができるストレージサブシステム及びストレージサブシステムの制御方法を提供することにある。本発明の一つの目的は、複数ユーザに資源を柔軟に割り当てることができる、かつ、より少ない情報量で、各ユーザの管理範囲が互いに影響しないように管理可能なストレージサブシステム及びストレージサブシステムの制御方法を提供することにある。本発明の一つの目的は、キャッシュメモリの管理単位に応じて管理情報を構成することにより、キャッシュ資源を論理的に分割して使用可能なストレージサブシステム及びストレージサブシステムの制御方法を提供することにある。本発明の他の目的は、後述する実施形態の記載から明らかになるであろう。

【課題を解決するための手段】

【0008】

上記課題を解決すべく、本発明に従うストレージサブシステムは、上位装置とのデータ授受をそれぞれ制御する複数のチャネルアダプタと、論理的な記憶領域をそれぞれ提供する複数の記憶デバイス群と、各記憶デバイス群とのデータ授受をそれぞれ制御する複数のディスクアダプタと、各チャネルアダプタ及び各ディスクアダプタによりそれぞれ使用されるキャッシュメモリと、キャッシュメモリを論理的に分割して構成される複数のキャッシュ分割領域と、各キャッシュ分割領域を管理するための管理情報を記憶する制御メモリと、を備えている。そして、管理情報は、各キャッシュ分割領域毎にそれぞれ設けられる分割管理情報と、各キャッシュ分割領域の全体に適用される共通管理情報とから構成される。

【0009】

キャッシュメモリと制御メモリとは、それぞれ別々のメモリ基板として実装することもできるし、両者を混載したメモリ基板を用いてもよい。また、メモリのある領域をキャッシュメモリとして使用し、他の領域を制御メモリとして使用してもよい。キャッシュメモリは、上位装置と記憶デバイス群との間で授受されるデータを一時的に（あるいは長期間にわたって）記憶するために用いられる。キャッシュメモリは、複数のキャッシュ分割領域に論理的に分割される。各キャッシュ分割領域は、管理情報によってそれぞれ独立して管理される。各キャッシュ分割領域は、例えば、各チャネルアダプタ毎にそれぞれ設ける（割り当てる）ことができる。そして、各キャッシュ分割領域は、例えば、それぞれ異なるユーザにより使用可能である。管理情報は、分割管理情報と共通管理情報とから構成される。分割管理情報は、各キャッシュ分割領域毎にそれぞれ設けられる情報である。共通管理情報は、全てのキャッシュ分割領域に適用される情報である。分割管理情報と共通管理情報とから管理情報を構成することにより、制御メモリの記憶資源を効率的に使用して、複数のキャッシュ分割領域をそれぞれ個別に使用可能である。

【0010】

各分割管理情報及び共通管理情報は、キャッシュ管理単位の属性に基づいて設定可能である。キャッシュ管理単位としては、例えば、スロットやセグメント等を挙げることができる。1つのスロットは、少なくとも1つ以上のセグメントから構成可能であり、1つのセグメントは、例えば、16KB程度のデータサイズを有する。キャッシュ管理単位の属性としては、例えば、フリー状態、ダーティ状態、クリーン状態を挙げることができる。フリー状態とは、キャッシュ管理単位が未使用である状態を示す。ダーティ状態とは、キャッシュ管理単位に格納されているデータが記憶デバイスに書き込まれていない状態を示す。クリーン状態とは、キャッシュ管理単位に格納されているデータが記憶デバイスに書き込み済である状態を示す。キャッシュ管理単位の属性によっては、各キャッシュ分割領域毎にそれぞれ個別の管理情報を設ける方が好ましい場合がある。逆に、キャッシュ管理単位の属性によっては、全てのキャッシュ分割領域に適用可能な管理情報もある。

【0011】

例えば、管理情報が、複数種類のサブ管理情報から構成される場合は、一部のサブ管理情報を各キャッシュ分割領域にそれぞれ分割して設け、残りのサブ管理情報を全てのキャッシュ分割領域に適用させることができる。この場合、サブ管理情報のうちいずれを分割管理情報として使用し、いずれを共通管理情報として使用するかは、キャッシュ管理単位の属性に基づいて決定することができる。

【0012】

例えば、管理情報は、複数種類のキューと、これら各キューにそれぞれ関連付けられたカウンタとを含んで構成可能である。この場合、キャッシュ管理単位の属性に基づいて、各キュー及び各カウンタの一部を各キャッシュ分割領域毎にそれぞれ設けることにより各分割管理情報を構成することができる。そして、各キューの残り及び各カウンタの残りを共通管理情報として使用可能である。「キャッシュ管理単位の属性に基づいて」とは、「キュー（及びカウンタ）の属性に基づいて」と言い換えることも可能である。

【0013】

ここで、互いに関連付けられているキューとカウンタのうち、一方は分割管理情報を構成し、他方は共通管理情報として使用することもできる。即ち、例えば、互いに関連付けられているキューとカウンタのうち、カウンタのみを分割して各キャッシュ分割領域にそれぞれ割り当てることができる。

【0014】

さらに、各キューにはそれぞれキュー管理テーブルを関連付けることができる。そして、分割管理情報を構成するキューに関連付けられるキュー管理テーブルは、各キャッシュ分割領域毎にそれぞれ設けることができる。即ち、キューとキュー管理テーブルとは常に一体的に使用されるもので、キューが各キャッシュ分割領域毎にそれぞれ設けられる場合は、キュー管理テーブルも各キャッシュ分割領域にそれぞれ設けられる。

【0015】

管理情報は、未使用状態のキャッシュ管理単位が接続されるフリーキュー及びこのフリーキューに関連付けられるフリーキューカウンタと、記憶デバイス群に反映前のダーティ状態のデータを格納するキャッシュ管理単位が接続されるダーティキュー及びこのダーティキューに関連付けられるダーティキューカウンタと、記憶デバイス群に反映済のクリーン状態のデータを格納するキャッシュ管理単位が接続されるクリーンキュー及びこのクリーンキューに関連付けられるクリーンキューカウンタと、使用中のキャッシュ管理単位の総量をカウントする使用中カウンタとを含んで構成可能である。そして、フリーキューカウンタと、クリーンキューと、クリーンキューカウンタ及び使用中カウンタは、分割管理情報として、各キャッシュ分割領域毎にそれぞれ設けることができる。また、フリーキューと、ダーティキュー及びダーティキューカウンタは、共通管理情報として使用することができる。

【0016】

各キャッシュ分割領域のうち1つのキャッシュ分割領域を共用領域として設定し、この共用領域に属する資源を割り当てることにより、新たなキャッシュ分割領域を設定することもできる。例えば、初期状態では、全てのキャッシュ領域を共用領域として使用する。そして、新たにキャッシュ分割領域を生成する場合は、共用領域から一部のキャッシュ領域を切り取って、新しいキャッシュ分割領域に割り当てる。なお、キャッシュ分割領域を削除する場合は、このキャッシュ分割領域に割り当てられていたキャッシュ領域を、共用領域に戻せばよい。

【発明を実施するための最良の形態】**【0017】**

以下、図1～図18に基づき、本発明の実施の形態を説明する。本実施形態では、上位装置とのデータ授受をそれぞれ制御する複数の上位インターフェース制御部と、論理的な記憶領域をそれぞれ提供する複数の記憶デバイス群と、各記憶デバイス群とのデータ授受をそれぞれ制御する複数の下位インターフェース制御部と、各上位インターフェース制御

部及び各下位インターフェース制御部によりそれぞれ使用されるメモリ部と、を備えたストレージサブシステムが開示される。そして、本実施例では、メモリ部が提供するキャッシュ領域を複数のキャッシュ分割領域に分割する。さらに、本実施例では、メモリ部でデータを管理するためのキャッシュ管理単位の属性に応じて、管理情報を各キャッシュ分割領域毎にそれぞれ分割し、各管理情報に基づいて、各キャッシュ分割領域毎にそれぞれデータを管理させる。

【実施例 1】

【0018】

図 1 は、ストレージサブシステム 10 の外観構成を示す概略斜視図である。ストレージサブシステム 10 は、例えば、基本筐体 11 と複数の増設筐体 12 とから構成することができる。基本筐体 11 は、ストレージサブシステム 10 の最小構成単位であり、記憶機能及び制御機能の両方を備えている。増設筐体 12 は、ストレージサブシステム 10 のオプションであり、基本筐体 11 の有する制御機能により制御される。例えば、最大 4 個の増設筐体 12 を基本筐体 11 に接続可能である。

【0019】

基本筐体 11 には、複数の制御パッケージ 13 と、複数の電源ユニット 14 と、複数のバッテリーユニット 15 と、複数のディスクドライブ 80 とがそれぞれ着脱可能に設けられている。増設筐体 12 には、複数のディスクドライブ 80 と、複数の電源ユニット 14 及び複数のバッテリーユニット 15 が着脱可能に設けられている。また、基本筐体 11 及び各増設筐体 12 には、複数の冷却ファン 16 がそれぞれ設けられている。

【0020】

制御パッケージ 13 は、後述するチャネルアダプタ 20、ディスクアダプタ 30 及びキャッシュメモリ 40 等をそれぞれ実現するためのモジュールである。即ち、基本筐体 11 には、複数のチャネルアダプタパッケージ、複数のディスクアダプタパッケージ及び 1 つ以上のメモリパッケージがそれぞれ着脱可能に設けられ、パッケージ単位で交換可能となっている。

【0021】

図 2 は、ストレージサブシステム 10 を含む記憶システムの全体概要を示すブロック図である。ストレージサブシステム 10 は、通信ネットワーク CN1 を介して、複数のホストコンピュータ 1A ~ 1C（以下、特に区別しない場合は、「ホストコンピュータ 1」と呼ぶ）と双方向通信可能にそれぞれ接続されている。ここで、通信ネットワーク CN1 は、例えば、LAN (Local Area Network)、SAN (Storage Area Network)、インターネットあるいは専用回線等である。LAN を用いる場合、ホストコンピュータ 1 とストレージサブシステム 10 との間のデータ転送は、TCP/IP (Transmission Control Protocol/Internet Protocol) プロトコルに従って行われる。SAN を用いる場合、ホストコンピュータ 1 とストレージサブシステム 10 とは、ファイバチャネルプロトコルに従ってデータ転送を行う。また、ホストコンピュータ 1 がメインフレームの場合は、例えば、FICON (Fibre Connection: 登録商標)、ESCON (Enterprise System Connection: 登録商標)、ACONARC (Advanced Connection Architecture: 登録商標)、FIBARC (Fibre Connection Architecture: 登録商標) 等の通信プロトコルに従ってデータ転送が行われる。

【0022】

各ホストコンピュータ 1 は、例えば、サーバ、パーソナルコンピュータ、ワークステーション、メインフレーム等として実現されるものである。例えば、各ホストコンピュータ 1 は、図外に位置する複数のクライアント端末と別の通信ネットワークを介して接続されている。各ホストコンピュータ 1 は、例えば、各クライアント端末からの要求に応じて、ストレージサブシステム 10 にデータの読み書きを行うことにより、各クライアント端末へのサービスを提供する。図中では、1 つのみ図示してあるが、ストレージサブシステム

1 0 内には、複数の仮想筐体 (SLPR: Storage Logical Partition) を設定可能である。

【 0 0 2 3 】

SLPRとは、ストレージサブシステム 1 0 内の各種物理的資源及び論理的資源を各ユーザ毎に (あるいは各アプリケーションプログラム毎に) 分割して割り当てることにより構成された領域である。即ち、例えば、SLPRには、それぞれ専用のチャンネルアダプタ 2 0 と、それぞれ専用のキャッシュ領域 4 0 と、それぞれ専用の仮想的な論理デバイス (VDEV) 7 0 等とを備えている。即ち、SLPRは、仮想的な小型のストレージサブシステムであるかのように振る舞う。

【 0 0 2 4 】

SLPRには、少なくとも 1 つ以上のキャッシュ分割領域 (CLPR: Cache Logical Partiton) を設けることができる。図 2 中では、左側に SLPR を 1 つだけ示してあるが、複数の SLPR を設けることができる。そして、1 つの SLPR 内には、1 つまたは複数の CLPR を設定することができる。

【 0 0 2 5 】

CLPRとは、キャッシュメモリ 4 0 を論理的に複数の領域に分割したものである。CLPRは、各チャンネルアダプタ 2 0 毎にそれぞれ設定可能である。例えば、チャンネルアダプタ 2 0 を n 個実装する場合は、n 個の CLPR を設けることができる。例えば、n = 3 2 に設定可能であるが、本発明はこれに限定されない。各 CLPR0 ~ CLPRn は、それぞれ互いに独立して使用されるものであり、各ホストコンピュータ 1 は、それぞれが利用可能な CLPR のみを独占的に使用することができる。そして、ある CLPR に対するホストコンピュータ 1 からの操作は、他の CLPR にできるだけ影響を与えないように構成されている。即ち、特定のホストコンピュータ 1 からのアクセスが集中した場合でも、そのホストコンピュータ 1 が利用可能な CLPR のみで必要なキャッシュ容量が確保され、他の CLPR の資源 (キャッシュ領域) を奪うことがないように構成されている。

【 0 0 2 6 】

なお、図中右側に示す CLPR0 は、SLPR が定義されていない共用領域であり、この共用領域 CLPR0 には、ストレージサブシステム 1 0 全体で使用される各種の情報が記憶される。また、後述のように、初期状態では、キャッシュメモリ 4 0 の全領域が共用領域 CLPR0 に割り当てられている。そして、共用領域 CLPR0 からキャッシュ領域を所定量だけ切り出すことにより、新たな CLPR を設定するようになっている。

【 0 0 2 7 】

図 2 に示す例では、ホストコンピュータ 1 A は、共用領域 CLPR0 のみにアクセスして、データの入出力を行うことができる。ホストコンピュータ 1 B は CLPR1 のみに、ホストコンピュータ 1 C は CLPR2 のみに、ホストコンピュータ 1 N は CLPRn のみに、それぞれアクセス可能であり、他の CLPR の利用や参照等を行うことはできない。

【 0 0 2 8 】

SVP (Service

Processor) 9 0 は、ストレージサブシステム 1 0 の管理や監視を行うためのコンピュータ装置であり、管理用のサーバ機能を提供する。SVP 9 0 は、例えば、ストレージサブシステム 1 0 内に設けられた LAN 等の内部ネットワーク CN 3 (図 3 参照) を介して、各チャンネルアダプタ 2 0 や各ディスクアダプタ 3 0 等から各種の環境情報や性能情報等を収集する。SVP 9 0 が収集する情報としては、例えば、装置構成、電源アラーム、温度アラーム、入出力速度 (IOPS) 等が挙げられる。SVP 9 0 と各管理端末 2 A ~ 2 N, 3 とは、例えば、LAN 等の通信ネットワーク CN 2 を介して接続されている。管理者は、管理端末を介して SVP 9 0 にログインすることにより、権限のある範囲内において、例えば、RAID 構成の設定、各種パッケージ (チャンネルアダプタパッケージ、ディスクアダプタパッケージ、メモリパッケージ、ディスクドライブ等) の閉塞処理や各種設定変更等を行うことができる。

【 0 0 2 9 】

SVP 9 0 には、複数の管理端末 2 A ~ 2 N, 3 を接続可能である。ここで、管理端末 2 A ~ 2 N は、各 SLPR 毎に設けられる端末であり、管理端末 3 は、ストレージサブシステム 1 0 の全体を管理するために設けられる端末である。以下の説明では、1 つの SLPR に 1 つの CLPR を設ける場合を例に挙げて説明する。従って、管理端末 2 A ~ 2 N は、各 CLPR をそれぞれ管理する管理者（以下、分割管理者と呼ぶ）によりそれぞれ操作されるサブ端末である。管理端末 3 は、ストレージサブシステム 1 0 の全体を管理するシステム管理者（あるいは全体管理者と呼ぶこともできる）により操作される全体端末である。

【 0 0 3 0 】

各サブ端末 2 A ~ 2 N の分割管理者は、自己が管理権限を有する CLPR についてのみ各種設定変更等を行うことができ、他の CLPR の構成等を参照したり変更することは許可されない。これに対し、システム管理者は、各 CLPR を含めてストレージサブシステム 1 0 の全体の各種設定変更等を行うことができる。

【 0 0 3 1 】

システム管理者は、全体端末 3 を介して SVP 9 0 にログインし、ストレージサブシステム 1 0 の有する物理的資源及び論理的資源を適宜分割することにより、各ユーザ毎に SLPR (CLPR) を設定することができる。また、システム管理者は、各分割管理者に対して、ユーザ ID 等を発行することもできる。分割管理者は、システム管理者から発行された専用のユーザ ID を用いて SVP 9 0 にログインすることができる。分割管理者は、サブ端末 2 を操作することにより、自己の管理下にある CLPR 内の設定を変更することができる。

【 0 0 3 2 】

図 3 は、ストレージサブシステム 1 0 の論理的構成に着目したブロック図である。ストレージサブシステム 1 0 は、複数のチャネルアダプタ（以下、CHA）2 0 と、複数のディスクアダプタ（以下、DKA）3 0 と、少なくとも 1 つ以上のキャッシュメモリ 4 0 及び共有メモリ 5 0 と、スイッチ部 6 0 と、複数の仮想的論理デバイス（VDEV）7 0 と、複数のディスクドライブ 8 0 と、SVP 9 0 とを備えている。

【 0 0 3 3 】

各 CHA 2 0 は、各ホストコンピュータ 1 との間のデータ転送を制御するもので、通信ポート 2 1 を備えている。ストレージサブシステム 1 0 には、例えば 3 2 個の CHA 2 0 を設けることができる。CHA 2 0 は、例えば、オープン系用 CHA、メインフレーム系用 CHA 等のように、ホストコンピュータ 1 の種類に応じて用意される。

【 0 0 3 4 】

各 CHA 2 0 は、それぞれに接続されたホストコンピュータ 1 から、データの読み書きを要求するコマンド及びデータを受信し、ホストコンピュータ 1 から受信したコマンドに従って動作する。DKA 3 0 の動作も含めて先に説明すると、例えば、CHA 2 0 は、ホストコンピュータ 1 からデータの読出し要求を受信すると、読出しコマンドを共有メモリ 5 0 に記憶させる。DKA 3 0 は、共有メモリ 5 0 を随時参照しており、未処理の読出しコマンドを発見すると、ディスクドライブ 8 0 からデータを読み出して、キャッシュメモリ 4 0 に記憶させる。CHA 2 0 は、キャッシュメモリ 4 0 に移されたデータを読み出し、コマンド発行元のホストコンピュータ 1 に送信する。

【 0 0 3 5 】

CHA 2 0 は、ホストコンピュータ 1 からデータの書込み要求を受信すると、書込みコマンドを共有メモリ 5 0 に記憶させる。また、CHA 2 0 は、受信したデータ（ユーザデータ）をキャッシュメモリ 4 0 に記憶させる。ここで、ホストコンピュータ 1 から書込みを要求されたデータは、ディスクドライブ 8 0 に書き込まれていない「ダーティデータ」であるため、例えば複数箇所にそれぞれ記憶されて多重化される。

【 0 0 3 6 】

CHA 2 0 は、キャッシュメモリ 4 0 にデータを記憶した後、ホストコンピュータ 1 に対して書込み完了を報告する。そして、DKA 3 0 は、共有メモリ 5 0 に記憶された書込みコマンドに従って、キャッシュメモリ 4 0 に記憶されたデータを読出し、所定のディスクドライブ 8 0 に記憶させる。ディスクドライブ 8 0 に書き込まれたデータは、「ダーティデ

ータ」から「グリーンデータ」に属性が変化し、キャッシュメモリ 4 0 による多重管理から解放される。なお、本明細書において、「ダーティデータ」とは、ディスクドライブ 8 0 に書き込まれていない状態のデータを意味する。また、「クリーンデータ」とは、ディスクドライブ 8 0 に書き込み済のデータを意味する。

【0 0 3 7】

各 DKA 3 0 は、ストレージサブシステム 1 0 内に例えば 4 個や 8 個等のように複数個設けることができる。各 DKA 3 0 は、各ディスクドライブ 8 0 との間のデータ通信を制御する。各 DKA 3 0 と各ディスクドライブ 8 0 とは、例えば、S A N 等の通信ネットワーク C N 4 を介して接続されており、ファイバチャネルプロトコルに従ってブロック単位のデータ転送を行う。各 DKA 3 0 は、ディスクドライブ 8 0 の状態を随時監視しており、この監視結果は内部ネットワーク C N 3 を介して SVP 9 0 に送信される。各 CHA 2 0 及び各 DKA 3 0 は、例えば、プロセッサやメモリ等が実装されたプリント基板と、メモリに格納された制御プログラム（いずれも不図示）とをそれぞれ備えており、これらのハードウェアとソフトウェアとの協働作業によって、それぞれ所定の機能を実現するようになっている。

【0 0 3 8】

キャッシュメモリ 4 0 は、例えば、ユーザデータ等を記憶するものである。キャッシュメモリ 4 0 は、例えば不揮発メモリから構成される。キャッシュメモリ 4 0 は、複数のメモリから構成することができ、ダーティデータを多重管理することができる。本実施例では、キャッシュメモリ 4 0 が有する全キャッシュ領域を複数個に分割することにより、各 CLPR0~n を設定している。

【0 0 3 9】

共有メモリ（あるいは制御メモリ）5 0 は、例えば不揮発メモリから構成される。共有メモリ 5 0 には、例えば、制御情報や管理情報等が記憶される。これらの制御情報等の情報は、複数の共有メモリ 5 0 により多重管理することができる。共有メモリ 5 0 及びキャッシュメモリ 4 0 は、それぞれ複数個設けることができる。また、同一のメモリ基板にキャッシュメモリ 4 0 と共有メモリ 5 0 とを混在させて実装することもできる。あるいは、メモリの一部をキャッシュ領域として使用し、他の一部を制御領域として使用することもできる。

【0 0 4 0】

スイッチ部 6 0 は、各 CHA 2 0 と、各 DKA 3 0 と、キャッシュメモリ 4 0 と、共有メモリ 5 0 とをそれぞれ接続するものである。これにより、全ての CHA 2 0、DKA 3 0 は、キャッシュメモリ 4 0 及び共有メモリ 5 0 にそれぞれアクセス可能である。スイッチ部 6 0 は、例えば超高速クロスバススイッチ等として構成することができる。

【0 0 4 1】

ストレージサブシステム 1 0 は、多数のディスクドライブ 8 0 を実装可能である。各ディスクドライブ 8 0 は、例えば、ハードディスクドライブ（H D D）や半導体メモリ装置等として実現可能である。ディスクドライブ 8 0 は、物理的な記憶デバイス（PDEV）である。そして、RAID 構成等によっても相違するが、例えば、4 個 1 組のディスクドライブ 8 0 が提供する物理的な記憶領域上には、仮想的な論理デバイス（VDEV）7 0 が構築される。なお、VDEV 7 0 上にさらに仮想的な論理デバイスを設定可能である。また、ストレージサブシステム 1 0 により使用される記憶資源は、全てストレージサブシステム 1 0 内に設ける必要はなく、ストレージサブシステム 1 0 の外部に存在する記憶資源を利用することもできる。例えば、外部の他社製ストレージサブシステムが有する記憶デバイスを特定の VDEV 7 0 に割り付け、他社製記憶デバイスへのアクセスパス等を管理することにより、他社の記憶デバイスをあたかも自己の記憶デバイスであるかのように使用可能である。

【0 0 4 2】

図 4 は、各 CLPR が資源を分割して利用する様子を模式的に示す説明図である。図中では、共用領域 CLPR0 と 2 個の専用 CLPR1、2 とによって、キャッシュメモリ 4 0 を分割して使用する場合を示してある。各 CLPR には、それぞれキャッシュ領域が割り当てられている。各 CLPR がそれぞれ使用可能なキャッシュ領域の最大値（最大割当量）は、システム管理者

により設定される。各CLPRにそれぞれ割り当てられたキャッシュ領域には、各CLPRを利用するホストコンピュータ 1 により使用されるデータ（ライトデータ、リードデータ）が格納される。図 4 中では、データが格納された領域に斜線を付し、「使用中」と表記してある。各CLPRは、それぞれが使用可能な最大割当量までのキャッシュ領域を使用可能であり、他のCLPRのキャッシュ領域を使用することはできない。

【0 0 4 3】

さらに、各CLPRは、それぞれのキャッシュ領域に加えて、他のキャッシュ関連データを管理している。キャッシュ関連データとしては、例えば、DCR（F 1）と、サイドファイル F 2 と、PCR（F 3）とを挙げることができる。

【0 0 4 4】

DCR (Dynamic

Cache Residency) 1 0 0 とは、キャッシュ常駐化機能であり、VDEV 7 0 の特定の領域に存在するデータをキャッシュ領域に常駐化させるものである。これにより、重要なデータ群に対するアクセス性能を高めることができる。

【0 0 4 5】

サイドファイル F 2 とは、図外のリモートサイトにデータを転送してコピーさせるためのデータを記憶するファイルである。例えば、リモートコピー対象のデータがサイドファイル F 2 に所定量以上蓄積された場合は、通信ネットワークを介して、距離的に離れた場所にある別のストレージサブシステムにデータが転送され、保持される。リモートコピーの初期コピーでは、指定されたデータ群を丸ごとリモートサイトにコピーし、それ以後に更新されたデータは、差分ファイルとしてリモートコピーされる。

【0 0 4 6】

PCR (Partical

Cache Residence) 1 0 2 とは、例えば、ストレージサブシステム 1 0 にNAS (Network Attached Storage) 機能を設ける場合に使用されるもので、データの種別毎にキャッシュ領域に常駐化させる機能である。DCR（F 1）は、特定の記憶空間に存在するデータを常駐化させるものであるのに対し、PCR（F 3）は、特定種類のデータを常駐化させるものである。なお、これらキャッシュ関連データDCR（F 1）、サイドファイル F 2 及びPCR（F 3）は、それを必要とするCLPRのみに設けられる。即ち、常駐化機能が設定されていないCLPRには、キャッシュ常駐化に関するデータは記憶されない。同様に、リモートコピーが設定されていないCLPRには、サイドファイル F 2 は記憶されない。

【0 0 4 7】

そして、図 4 に示すように、各CLPRには、少なくとも 1 つ以上のVDEV 7 0 がそれぞれ割り当てられている。各CLPRのキャッシュ領域に書き込まれたデータは、それぞれに割り当てられたVDEV 7 0 の所定領域に書き込まれる。また、VDEV 7 0 から読み出されたデータは、対応するCLPRのキャッシュ領域に保持される。

【0 0 4 8】

図 5 は、キャッシュ領域に記憶されているデータを管理するための方法を示す説明図である。ストレージサブシステム 1 0 では、キャッシュメモリ 4 0 に記憶されているデータを効率的に検索するために、以下のような階層構造で管理している。

【0 0 4 9】

まず、ホストコンピュータ 1 からデータの入出力要求があると、この入出力要求に含まれるLBA (Logical Block Address) に基づいて、VDEV SLOT 番号 (VDEV SLOT #) を求める。そして、VDEV SLOT 番号に基づいて、VDSLOT-PAGE テーブル T 1 を参照することにより、次の階層へのポインタを取得する。VDSLOT-PAGE テーブル T 1 には、PAGE-DIR テーブル T 2 へのポインタが含まれている。PAGE-DIR テーブル T 2 には、PAGE-GRPP テーブル T 3 へのポインタが含まれている。さらに、PAGE-GRPP テーブル T 3 には、GRPT 1 テーブル T 4 へのポインタが含まれている。GRPT 1 テーブル T 4 には、GRPT 2 テーブル T 5 へのポインタが含まれている。GRPT 2 テーブル T 5 には、SLCB (スロット制御テーブル) T 6 へのポインタが含まれ

ている。

【0050】

このように、LBAに基づいて、各テーブルT1～T5を順番に参照していくことにより、SLCBテーブルT6に到達する。SLCBテーブルT6には、少なくとも1つ以上のSGCB（セグメント制御ブロック）テーブルT7が関連付けられている。SGCBテーブルT7は、キャッシュ管理の最小単位であるセグメントに関する制御情報を格納している。1つのスロットには、1～4個のセグメントを関連付けることができる。1つのセグメントには、例えば、48KBまたは64KBのデータを格納することができる。

【0051】

キャッシュ管理の最小単位はセグメントであるが、ダーティデータ（物理ディスクへの書き込み前の状態）と、クリーンデータ（物理ディスクへの書き込み後の状態）の各状態間の遷移は、スロット単位で行われる。また、キャッシュ領域の確保（リザーブ）及び解放（リリース）も、スロット単位またはセグメント単位で行われる。

【0052】

SLCBテーブルT6には、図8に示すように、後方ポインタと、前方ポインタと、キューステータスと、VDSLOT#と、スロットステータスと、CLPR情報と、少なくとも1つ以上のSGCBポインタとを含めて構成することができる。キューステータスは、そのSLCBテーブルT6に関連付けられているキューの種別やキュー番号を含む。スロットステータスは、そのSLCBテーブルT6に対応するスロットの状態を含む。CLPR情報は、そのSLCBテーブルT6が属するCLPRに関する情報（CLPR番号等）を含む。SGCBポインタは、そのSLCBテーブルT6に関連付けられているSGCBテーブルT7を特定するためのポインタが含まれている。

【0053】

なお、SLCBテーブルT6及びSGCBテーブルT7は、例えば、ポインタ変更等の競合が発生して整合性が失われないように、VDEV70単位で排他制御が行われる。即ち、SLCBテーブルT6またはSGCBテーブルT7が属するVDEV70について、ロックを取得した場合にのみ、SLCBテーブルT6またはSGCBテーブルT7を操作できるようになっている。本明細書では、このVDEV単位の排他制御を「VDEVロック」と呼ぶ。VDEVロックは、論理ボリューム単位の排他制御であると言い換えることもできる。

【0054】

図6は、SLCBテーブルT6の状態遷移を示す説明図である。SLCBテーブルT6は、例えば、未使用状態ST1、デバイス反映済み状態ST2、パリティ無し&デバイス反映前状態ST3、パリティ生成中状態ST4及びパリティ有り&ドライブ反映前状態ST5の5つのステータスを取ることができる。

【0055】

未使用状態ST1とは、そのスロットが未使用であることを示す状態である。未使用状態にあるスロット（SLCB）は、フリーSGCBキュー（以下、フリーキューと略記）により管理される。デバイス反映済み状態ST2とは、そのスロットに格納されているデータが物理デバイス（ディスクドライブ80）に書き込み済みであることを示す状態である。デバイス反映済み状態ST2にあるスロットは、クリーンキューにより管理される。

【0056】

パリティ無し&デバイス反映前状態ST3とは、例えばRAID5等のようにパリティデータを必要とする場合において、そのスロットに格納されているデータについて、未だパリティデータが生成されていないことを示す状態である。パリティ無し&デバイス反映前状態ST3にあるスロットは、中間ダーティキューにより管理される。パリティ生成中状態ST4とは、そのスロットに格納されているデータについて、パリティデータを生成している最中であることを示す状態である。パリティ生成中状態ST4にあるスロットは、ホストダーティキューにより管理される。

【0057】

パリティ有り&ドライブ反映前状態ST5とは、そのスロットに格納されているデータについて、パリティデータが生成されているが、物理デバイスに未だ書き込まれてい

ないことを示す状態である。パリティ有り&ドライブ反映前状態 S T 5 にあるスロットは、物理ダーティキュー（以下、ダーティキューと略記）により管理される。

【 0 0 5 8 】

ホストコンピュータ 1 からライトアクセスがあった場合、S T 1 から S T 2 を経て S T 3 に状態が遷移し、パリティデータの生成が開始される。そして、書込みを要求されたデータについてパリティデータが生成されると、状態は S T 3 から S T 4 を経て S T 5 に変化する。ホストコンピュータ 1 から CLPR のキャッシュ領域の 1 つまたは複数のスロットにデータが書き込まれた時点で、ホストコンピュータ 1 に対して書込み完了報告が送信される。この時点では、書込みを要求されたデータは、キャッシュ領域上にのみ存在するので、ダーティ状態（S T 5）となっている。

【 0 0 5 9 】

そして、ダーティ状態にあるスロットのデータが所定の物理デバイスに書き込まれると（デステージ）、そのスロットは、ダーティ状態からクリーン状態に変化する（S T 5 → S T 2）。クリーン状態にあるデータは、物理デバイスに書き込まれているので、もしも必要ならばキャッシュ領域から消去しても構わない。クリーン状態にあるスロットに格納されているデータをキャッシュ上から消去すると、そのスロットは、クリーン状態から未使用状態へ戻る（S T 2 → S T 1）。

【 0 0 6 0 】

ホストコンピュータ 1 からリードアクセスがあった場合は、図 5 と共に述べた各階層テーブル T 1 ～ T 6 を順番に辿ることにより、要求されたデータが特定される。そして、ホストコンピュータ 1 が要求しているデータを物理デバイスから読み出して、キャッシュ領域の 1 つまたは複数のスロットに格納させる。物理デバイスから読み出されたデータを格納するスロットは、クリーン状態を有する。

【 0 0 6 1 】

図 7 は、キュー構造を示す説明図である。キュー 1 0 0 は、同一状態にある SLCB テーブル T 6 を一方向または双方向につなげる待ち行列として構成される。SLCB テーブル T 6 の状態が変化すると、例えば、FIFO（First-In First-Out：先入れ先出し）や LRU（Least Recently Used：最長未使用時間）等のような予め設定されたアルゴリズムに基づいて、キュー 1 0 0 の構成は変化する。

【 0 0 6 2 】

SLCB テーブル T 6 への操作が競合して整合性が失われるのを防止するために、各キュー 1 0 0 毎に排他制御を行う。即ち、操作対象のキューのロックを取得した場合にのみ、そのキューへの操作を行うことができる。本明細書では、このキュー単位の排他制御を「キューロック」と呼ぶ。従って、上述した VDEV ロックとキューロックとの 2 段階の排他制御の下で、SLCB テーブル T 6 は操作可能となっている。つまり、操作しようとする SLCB テーブル T 6 が属する VDEV 7 0 のロックと、その SLCB テーブル T 6 が接続されているキューのロックとの 2 つのロックを取得した場合のみ、その SLCB テーブル T 6 の内容を変更することができる。

【 0 0 6 3 】

このように、キュー 1 0 0 への操作は排他制御されるため、ロック取得時のオーバーヘッドを考慮して、同一種類のキューを複数本設けている。クリーンキュー、ダーティキュー、フリーキュー等は、それぞれ複数本のキューから構成される。

【 0 0 6 4 】

キューセットを構成する個々のキューには、それぞれカウンタ 2 0 0 が関連付けられている。カウンタ 2 0 0 は、そのキューに属する SLCB テーブル T 6 及び SGCB テーブル T 7 の数を管理するものである。カウンタ 2 0 0 に管理されている SLCB 数等に基づいて、後述の過負荷制御等を行うことができる。

【 0 0 6 5 】

また、キューセットを構成する個々のキュー 1 0 0 には、キュー管理テーブル 3 0 0 も

関連付けられる。キュー管理テーブル 3 0 0 は、キュー 1 0 0 の構造等を管理するものである。キュー 1 0 0、カウンタ 2 0 0 及びキュー管理テーブル 3 0 0 は、それぞれ共有メモリ 5 0 に記憶される。

【 0 0 6 6 】

図 9 は、各 CLPR のキャッシュ領域の使用状態と各キューとの関係を模式的に示す説明図である。各 CLPR のキャッシュ領域は、その使用状態に応じて、「使用中領域」と「未使用領域」とに分けることができる。使用中領域とは、データを格納しているスロット群（セグメント群）を意味する。未使用領域とは、空いているスロット群（セグメント群）を意味する。

【 0 0 6 7 】

使用されているセグメントの総数と 1 つのセグメントあたりのデータサイズとを乗算したものが、その CLPR における使用中領域のサイズとなる。使用されているセグメントには、クリーン状態にあるセグメントとダーティ状態にあるセグメントとが含まれる。従って、その CLPR に割り当てられているキャッシュ領域のうち現在使用されている領域のサイズは、（クリーン状態のセグメント数＋ダーティ状態のセグメント数）で表すことができる。未使用領域のサイズは、フリー状態にあるセグメント数で表すことができる。

【 0 0 6 8 】

フリーキュー 1 0 1 は、フリー状態（未使用状態）にある SLCB テーブル T 6 を管理するものである。フリーキュー 1 0 1 に接続されている SLCB テーブル T 6（及び SGCB テーブル T 7）の数は、フリーキュー 1 0 1 に関連付けられているフリーカウンタ 2 0 1 により把握されている。フリーキュー 1 0 1 の制御単位は VDEV であり、キューロックの単位も VDEV 単位である。そして、フリーキュー 1 0 1 は、例えば、FIFO に従って使用される。フリーキュー 1 0 1 は、例えば、VDEV 7 0 の実装数の 2 倍設けている。従って、実装可能な VDEV 7 0 の最大数を 5 1 2 とすると、フリーキュー 1 0 1 の総数は 1 0 2 4 となる。

【 0 0 6 9 】

クリーンキュー 1 0 2 は、クリーン状態にある SLCB 管理テーブル T 6 を管理するものである。クリーンキュー 1 0 2 に接続されている SLCB テーブル T 6（及び SGCB テーブル T 7）の数は、クリーンキュー 1 0 2 に関連付けられているクリーンカウンタ 2 0 2 によって把握されている。クリーンキュー 1 0 2 の制御単位は VDEV であり、キューロックの単位も VDEV 単位である。クリーンキュー 1 0 2 は、例えば、LRU に従って使用される。クリーンキュー 1 0 2 は、例えば、VDEV 7 0 の実装数の 4 倍設けている（ $512 \times 4 = 2048$ ）。

【 0 0 7 0 】

ダーティキュー 1 0 3 は、ダーティ状態にある SLCB 管理テーブル T 6 を管理するものである。ダーティキュー 1 0 3 に接続されている SLCB テーブル T 6（及び SGCB テーブル T 7）の数は、ダーティキュー 1 0 3 に関連付けられているダーティカウンタ 2 0 3 によって把握されている。ダーティキュー 1 0 3 の制御単位は物理ドライブである。ダーティキュー 1 0 3 は、例えば、ディスクドライブ 8 0 の実装数の 3 2 倍設けている。ディスクドライブ 8 0 を 2 0 4 8 個実装する場合、ダーティキュー 1 0 3 は 6 5 5 3 6 個設けることができる。ダーティキュー 1 0 3 は、物理デバイスに書込み前のデータ（セグメント）を管理するためのものである。従って、論理デバイス（VDEV 7 0）よりも物理デバイス（ディスクドライブ 8 0）の方に関連が深い。そこで、ダーティキュー 1 0 3 は、物理デバイスの数に依存して設定するようになっている。

【 0 0 7 1 】

使用中カウンタ 2 0 6 は、各 CLPR における使用中のセグメント数（ダーティ状態のセグメント数＋クリーン状態のセグメント数）をカウントするものである。

【 0 0 7 2 】

図 1 0 は、各 CLPR と各キュー及びカウンタ等の関係を示す説明図である。上述した各キュー、カウンタ及び管理テーブルは、キャッシュ管理単位（セグメントまたはスロット）を管理するための管理情報を構成する。本実施例では、セグメントまたはスロットの性質

(ステータス)に応じて、各CLPR毎にそれぞれ設ける管理情報(分割管理情報)と、全CLPRに適用する共通管理情報とに区別する。

【0073】

図10に示すように、各CLPRには、クリーンキュー102と、クリーンカウンタ(正確にはクリーンキューカウンタ)202と、フリーキューカウンタ201と、色分けキューカウンタ204と、使用中カウンタ206と、BINDキューカウンタ205とが、それぞれ設けられている。また、図示を省略しているが、クリーンキュー102の構成を管理するためのクリーンキュー管理テーブルも各CLPR毎にそれぞれ設けられている。以上のように、各CLPR毎にそれぞれ設けられる管理用の情報が分割管理情報を構成する。

【0074】

これに対し、フリーキュー101と、ダーティキュー103と、ダーティカウンタ(正確にはダーティキューカウンタ)203と、色分けキュー104と、BINDキュー105と、これら各キュー101, 103~105を管理するためのキュー管理テーブル301とは、全てのCLPRに共通に適用される共通管理情報を構成する。

【0075】

色分けキュー104とは、PCR機能によってキャッシュ領域に常駐化されるデータを管理するためのキューである。色分けキューカウンタ204は、PCRにより常駐化されているセグメント数をカウントする。

【0076】

BINDキュー105とは、DCR機能によってキャッシュ領域に常駐化されるデータを管理するためのキューである。BINDキューカウンタ205は、DCRにより常駐化されているセグメント数(またはスロット数)をカウントする。

【0077】

常駐化されるデータを格納するセグメントを管理するための色分けキュー104及びBINDキュー105をそれぞれ設けることにより、各CLPR内において、常駐化すべきデータを格納するセグメントがフリー状態に変更されるのを防止することができる。

【0078】

次に、管理情報を構成する各サブ管理情報(キュー、カウンタ、キュー管理テーブル)のうち、いずれを各CLPR毎にそれぞれ設け、いずれを全CLPRに適用するのかの理由を説明する。

【0079】

クリーンキュー102と、クリーンカウンタ202及びクリーンキュー管理テーブルは、それぞれ各CLPR毎に設けられている。フリー状態のセグメントが枯渇すると、クリーン状態のスロットが解放されてフリースロット化される。従って、クリーンキュー102は、最も使用頻度の高いキューである。もしも、全てのCLPRにおいて、1セットのクリーンキューを用いる場合は、あるCLPRにおいてフリーセグメントの枯渇が生じると、他のCLPRで利用されているデータを格納するクリーン状態のスロットが解放されてしまう可能性がある。即ち、一方のCLPRにおけるキャッシュ利用状態が他方のCLPRにおけるキャッシュ利用状態に影響を与えることになる。そこで、本実施例では、クリーンキュー102と、クリーンカウンタ202とを各CLPR毎にそれぞれ設けることができるようにしている。なお、キュー管理テーブルは、キューと一体的に取り扱われるものである。従って、クリーンキュー102を各CLPR毎にそれぞれ設ける場合、クリーンキュー管理テーブルも各CLPR毎にそれぞれ設けられる。

【0080】

使用中カウンタ206は、各CLPR毎にそれぞれ設けられる。使用中カウンタ206は、クリーン状態のセグメント数とダーティ状態のセグメント数とを加算した値をカウントするものである。従って、使用中カウンタ206のカウント値からクリーンカウンタ202のカウント値を減算して得られる値は、そのCLPRにおけるダーティ状態のセグメント数に他ならない。従って、クリーンカウンタ202及び使用中カウンタ206が各CLPR毎にそれぞれ設けられていれば、ダーティカウンタ203を各CLPR毎にそれぞれ設ける必要はな

い。

【0081】

ダーティキュー103は、物理デバイスに関連するキューであり、制御単位も物理デバイス単位となっている。従って、もしも、ダーティキュー103を各CLPR毎にそれぞれ設ける場合は、ダーティキュー管理テーブルを記憶するための記憶容量が増大し、共有メモリ50の記憶領域を圧迫する。例えば、キュー1本あたりに必要となる情報量は、キュー管理テーブルが16バイト、カウンタが4バイトである。キューの分割はキュー管理テーブルの分割を意味するため、共有メモリ50の多くの記憶領域が分割管理情報の記憶のために消費される。そこで、本実施例では、共有メモリ50の記憶資源を効率的に使用するために、ダーティキュー103及びダーティカウンタ203は、各CLPR毎に設けるのではなく、全てのCLPRについて適用することとした。

【0082】

また、共有メモリ50の記憶資源を有効利用する観点から、フリーキュー101も各CLPR毎にそれぞれ設けることなく、バイト数の少ないフリーカウンタ201のみを各CLPR毎にそれぞれ設けることとした。

【0083】

フリーキュー101と同様の理由により、色分けキュー104、BINDキュー105も、それぞれのカウンタ204、205のみを各CLPR毎にそれぞれ設け、キュー自体は全てのCLPRに適用することとした。

【0084】

このように、本実施例では、サブ管理情報の属性に応じて、即ち、サブ管理情報により管理されるキャッシュ管理単位の属性に応じて、サブ管理情報を分配する。従って、共有メモリ50の記憶資源を効率的に使用し、各CLPR間の独立性を担保できる。

【0085】

図11は、キャッシュ管理構造と排他制御の関係を示す説明図である。図11では、図10と共に述べた管理情報の一部のみが示されている。図11中の「VDEVロックGr」とは、VDEVロックのグループを示す。ディレクトリを操作するためにはVDEVロックの取得が必要となり、キューを操作するためにはキューロックの操作が必要となる。つまり、CLPRに管理されているデータを移動させるためには、そのデータを格納するセグメントを管理するキューへの操作権限と、そのセグメントが属するディレクトリへの操作権限との両方が必要となる。

【0086】

図11の右側に示すように、各CLPRには、1つまたは複数のVDEV70に関連付けることができる。あるCLPRに関連付けられているVDEV70を、別のCLPRに関連付けることもできる。図12は、CLPR0に関連付けられているVDEV#1をCLPR1へ付け替える様子を示す説明図である。

【0087】

VDEV70の所属先CLPRを変更する場合は、VDEV70とCLPRとの関連付けを変更し、さらに、移動対象のVDEV70に属するデータ群を新たなCLPRに関連付ける必要がある。即ち、移動対象のセグメント（またはスロット）を、移動元CLPRに関連付けられたキューから外し、移動先CLPRに関連付けられたキューに接続する。この操作に際して、VDEVロックとキューロックの両方が必要とされる。

【0088】

図13は、CLPRの構成を変更する場合の処理の一例を示す。図13には、移動元CLPRに関連付けられているVDEV70を、移動先CLPRに移動させる場合の概略が示されている。

【0089】

まず、管理端末3を介して、システム管理者がVDEV70の移動を要求すると（S1：YES）、CHA20やDKA30の代表プロセッサは、移動対象として指定されたVDEV70の固有情報を変更する（S2）。即ち、そのVDEV70が所属するCLPRに関する情報（帰属先CLPR情報）が、移動先のCLPRに関する情報に書き換えられる。

【0090】

次に、代表プロセッサは、移動対象のVDEV 7 0 に対応するSLCBテーブルT 6 を検索して抽出し、この移動対象となるSLCBテーブルT 6 について、CLPR情報を変更する（S 3）。即ち、図8と共に上述したように、移動対象となっている各SLCBテーブルT 6 について、その所属先CLPRに関する情報を移動先CLPRに関する情報に変更する。

【0091】

そして、代表プロセッサは、移動対象のSLCBテーブルT 6 のスロットステータスがクリーン状態であるか否かを検査する（S 4）。クリーン状態のスロットである場合（S4: YES）、即ち、移動対象のSLCBテーブルT 6 がクリーンキュー 1 0 2 に接続されている場合、代表プロセッサは、そのSLCBテーブルT 6 のキューステータスを、移動先CLPRに合わせて変更する（S 5）。つまり、移動先CLPRに設けられているの何番目のクリーンキュー 1 0 2 に接続するか等の情報を書き換える。

【0092】

代表プロセッサは、キューステータスを変更後、移動対象のSLCBテーブルT 6 を現在のクリーンキュー 1 0 2 から外す（S 6）。そして、代表プロセッサは、移動させるクリーンスロットの数に応じて、移動元CLPRに設けられている使用中カウンタ 2 0 6 及びクリーンカウンタ 2 0 2 のカウント値をそれぞれ減じさせる（S 7）。次に、代表プロセッサは、移動対象のSLCBテーブルT 6 を、移動先CLPRのクリーンキュー 1 0 2 に接続する（S 8）。そして、代表プロセッサは、移動先CLPRにおける使用中カウンタ 2 0 6 及びクリーンカウンタ 2 0 2 のカウント値をそれぞれ増加させて（S 9）、本処理を終了する。

【0093】

一方、移動対象のVDEV 7 0 に対応するSLCBテーブルT 6 がクリーン状態ではない場合（S4: NO）、代表プロセッサは、その移動対象のSLCBテーブルT 6 がフリー状態であるか否かを検査する（S 1 0）。移動対象のSLCBテーブルT 6 がフリー状態の場合（S10: YES）、代表プロセッサは、移動元及び移動先のフリーカウンタ 2 0 1 をそれぞれ変更する（S 1 1）。即ち、移動させるフリースロットの数に応じて、移動元のフリーカウンタ 2 0 1 を減算し、移動先のフリーカウンタ 2 0 1 を加算し、処理を終了する。移動対象のSLCBテーブルT 6 がダーティ状態の場合（S10: NO）、何もせずに処理を終了する。ダーティキュー 1 0 3 及びダーティカウンタ 2 0 3 は、全CLPRについて共通に適用されるので、キュー移動やカウンタの増減を行う必要はない。

【0094】

なお、図13中では省略しているが、色分けキュー 1 0 4 及びBINDキュー 1 0 5 に接続されるSLCBテーブルT 6 は、それぞれフリーキュー 1 0 1 と同様に処理される。色分けキュー 1 0 4 及びBINDキュー 1 0 5 は、各CLPR毎にそれぞれ分割して設けるのではなく、そのカウンタ 2 0 4, 2 0 5 のみを各CLPRに設けるようになっている。従って、キュー間の移動は生じず、移動元と移動先でのカウンタ値の変更のみが発生する。

【0095】

図14は、CLPRのキャッシュ割当量を変更する場合の概略処理を示すフローチャートである。システム管理者によるキャッシュ割当処理の変更操作が行われると（S31: YES）、代表プロセッサは、キャッシュ割当の変更に関連するSLCBテーブルT 6 を検出する（S 3 2）。そして、代表プロセッサは、変更に関連するSLCBテーブルT 6 のCLPR情報を、キャッシュ割当の変更に応じて、変更させる（S 3 3）。

【0096】

例えば、CLPR1のキャッシュ割当量を減らして、その分だけCLPR2のキャッシュ割当量を増加させる場合、この変動領域に位置するSLCBテーブルT 6 の帰属先CLPRは、CLPR1からCLPR2に変更される。そして、キャッシュ領域の割当変更に応じて、クリーンカウンタ 2 0 2 等の値を増減させて処理を終了する（S 3 4）。

【0097】

図15は、フリーセグメントを確保するためのフリーセグメント収集処理の概略を示すフローチャートである。ここで、フリーセグメントとは、フリー状態にあるセグメントを

意味する。本処理は、定期的に、あるいは、フリーセグメントの不足時（過負荷時）に、行うことができる。

【0098】

まず、クリーン状態のSLCBテーブルT6のうち最も使用頻度の低いものを検出し（S41）、検出されたSLCBテーブルT6のステータスを、クリーン状態からフリー状態に変更する（S42）。このステータス変更に伴って、SLCBテーブルT6は、クリーンキュー102からフリーキュー101に付け替えられる。また、クリーンカウンタ202及びフリーカウンタ201のカウント値もそれぞれ変更される。S41及びS42の処理は、所定量のフリーセグメントが確保されるまで、繰り返し行われる（S43）。

【0099】

図16は、CLPRを定義する様子を示す模式図である。本実施例では、CLPR0を共用のプール領域として使用する。CLPR0には、ストレージサブシステム10内で共通に使用される各種情報が記憶される。なお、上述のように、各種の管理情報は共有メモリ50に記憶されている。

【0100】

図16（a）に示すように、ユーザ専用のCLPRを1つも設定していない初期状態では、キャッシュメモリ40の有する全てのキャッシュ領域は、CLPR0に所属する。図16（b）に示すように、CLPR0に割り当てられたキャッシュ領域は、共通の諸情報等を記憶する「使用中領域」と「未使用領域」とに区別することができる。ユーザ専用のCLPR1を新たに設定する場合は、図16（c）に示すように、プール領域であるCLPR0から必要なだけのキャッシュ領域を切り取って、CLPR1に割り当てる。さらに、別のユーザ用にCLPR2を新たに設定する場合は、図16（d）に示すように、プール領域として使用されるCLPR0から必要な量だけキャッシュ領域を切り取り、新たに生成するCLPR2に割り当てる。ユーザ専用のCLPR1またはCLPR2を削除する場合は、削除されるCLPRに割り当てられているキャッシュ領域は、CLPR0に戻される。

【0101】

図17は、上述した処理を示すフローチャートである。初期状態では、キャッシュメモリ40の有する全てのキャッシュ領域が、プール領域として使用される共用領域CLPR0に割り当てられる（S51）。ストレージサブシステム10全体で使用する諸情報は、CLPR0に記憶され、運用が開始される（S52）。

【0102】

CLPRを新たに追加する場合は（S53：YES）、CLPRの構成変更を行う（S54）。このCLPR構成変更処理では、CLPR0からキャッシュ領域を切り取り、新たなCLPRに割り当てる等の処理を行う。一方、既に生成されたCLPRを削除する場合（S53：NO、S55：YES）、削除するCLPRに割り当てられていたキャッシュ領域をCLPR0に戻す等の処理を行う（S56）。なお、原則として、CLPR0を削除することはできない。

【0103】

図18は、各サブ端末2毎に、それぞれのサブ端末2が管理するCLPRに関する情報を表示等する様子を示す説明図である。

【0104】

SVP90は、認証部91と、構成管理部92とを備えている。各CLPRの管理者は、サブ端末2を介してSVP90にアクセスし、ユーザID等を入力する。認証部91によるユーザ認証の結果、正当なユーザであると認められた場合、CLPRの管理者は、自己の管理下にあるCLPRについて構成変更を行うことができる。この構成変更の操作は、構成管理部92によって各CLPRに反映される。また、構成管理部92から取得されたCLPRに関する構成情報は、サブ端末2に表示させることができる。

【0105】

サブ端末2に表示可能な情報としては、例えば、CLPRに割り当てられている最大キャッシュ量、CLPRに割り当てられているVDEVに関する情報（VDEV番号やボリュームサイズ等）、使用中のキャッシュサイズ、未使用のキャッシュサイズ、CLPRへのIO頻度等を挙げるこ

とができる。なお、ストレージサブシステム 1 0 の全体を管理するシステム管理者は、全体端末 3 を介して、全ての CLPR に関する情報を取得することができ、各 CLPR の構成変更等を行うことができる。

【0 1 0 6】

このように、本実施例によれば、ストレージサブシステム 1 0 内に設定した各 CLPR 毎に、それぞれ管理用の情報を分割する構成とした。従って、各 CLPR 間の干渉を抑制することができ、使い勝手を高めることができる。即ち、一方の CLPR に対するアクセスが増大した場合でも、他の CLPR に影響が及ぶのを極力防止することができる。

【0 1 0 7】

また、本実施例では、セグメントまたはスロットの属性に基づいて、管理情報を分配する構成とした。従って、全ての管理情報を単純に各 CLPR にそれぞれ設ける構成と比較して、管理情報全体のデータサイズを低減することができ、共有メモリ 5 0 を効率的に使用することができる。

【0 1 0 8】

なお、本発明は、上述した実施の形態に限定されない。当業者であれば、本発明の範囲内で、種々の追加や変更等を行うことができる。

【図面の簡単な説明】

【0 1 0 9】

【図 1】本発明の実施例に係わるストレージサブシステムの外観図である。

【図 2】ストレージサブシステムの論理的な概略構成を示すブロック図である。

【図 3】ストレージサブシステムのより詳細なブロック図である。

【図 4】CLPR 毎に割り当てられるキャッシュ領域と VDEV の関係を示す説明図である。

【図 5】階層構造を有するキャッシュ管理方法を示す説明図である。

【図 6】SLCB テーブルの状態遷移図である。

【図 7】キュー、カウンタ及びキュー管理テーブルの関係を示す説明図である。

【図 8】SLCB テーブルの概略構造を示す説明図である。

【図 9】キャッシュ領域と各キュー及びカウンタとの関係を示す説明図である。

【図 1 0】CLPR の管理情報を分配する様子を示す模式図である。

【図 1 1】VDEV ロック及びキューロックによってデータ移動が可能となる様子を示す説明図である。

【図 1 2】CLPR0 から CLPR1 に VDEV を移動させる様子を示す説明図である。

【図 1 3】VDEV を移動させる場合の処理を示すフローチャートである。

【図 1 4】CLPR へのキャッシュ割当を変更する場合のフローチャートである。

【図 1 5】フリーセグメントの収集処理を示すフローチャートである。

【図 1 6】プール領域として用いられる CLPR0 からキャッシュ領域を切り取って、新たな CLPR を設定する様子を示す説明図である。

【図 1 7】CLPR を定義する場合の処理を示すフローチャートである。

【図 1 8】各 CLPR に関する情報をサブ管理端末に表示させる様子を示す説明図である。

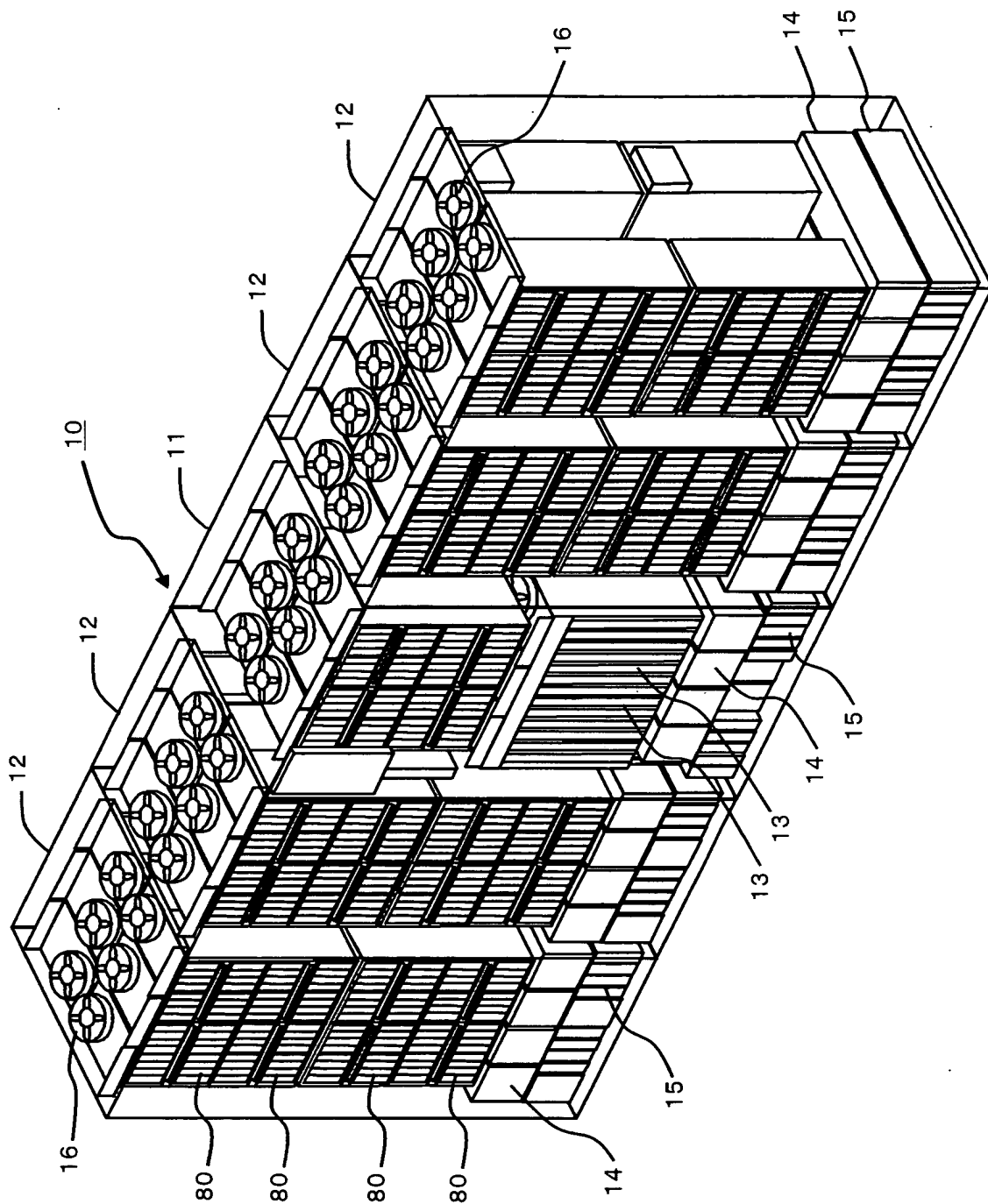
【符号の説明】

【0 1 1 0】

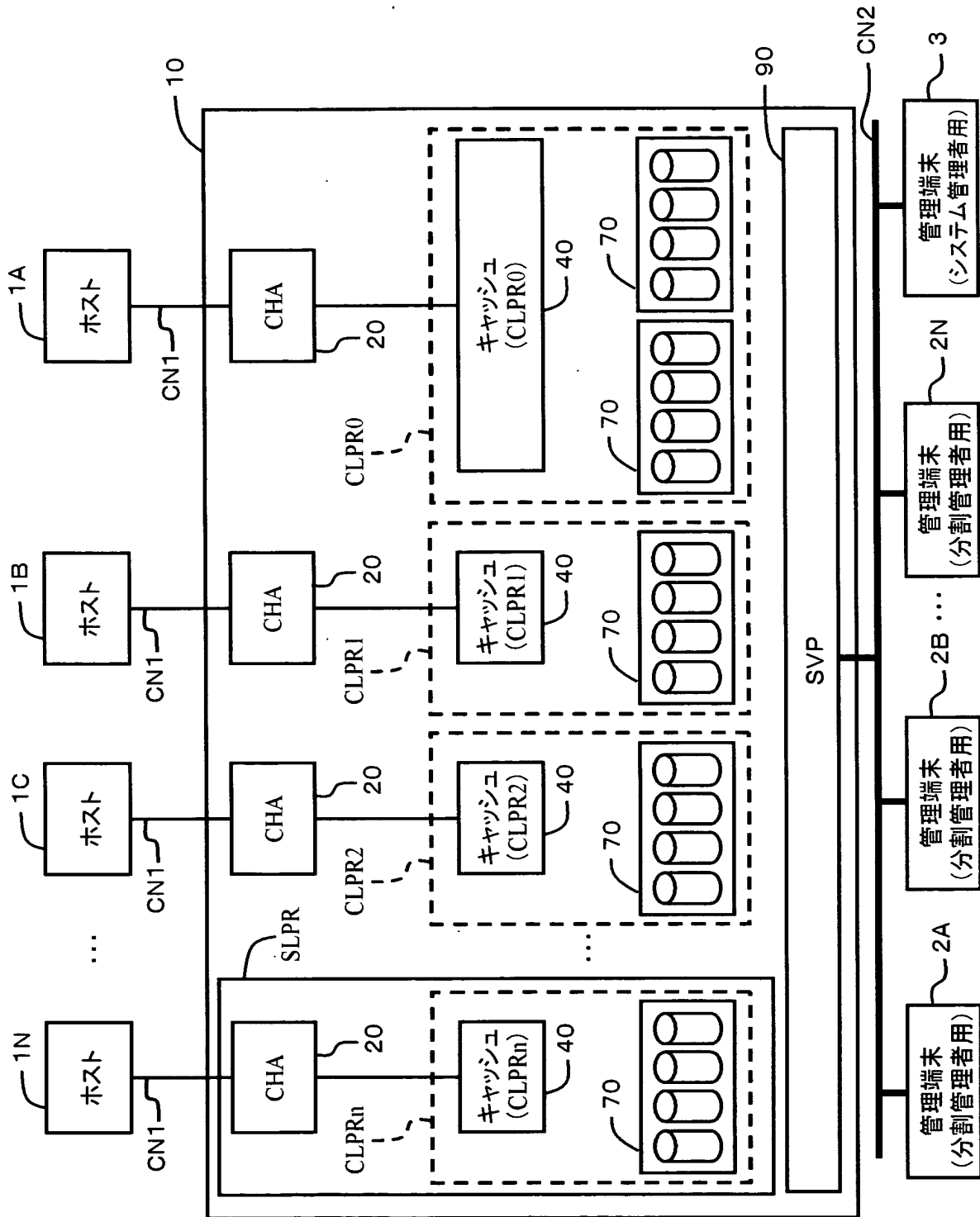
1…ホストコンピュータ、2…サブ端末、3…全体端末、1 0…ストレージサブシステム、1 1…基本筐体、1 2…増設筐体、1 3…制御パッケージ、1 4…電源ユニット、1 5…バッテリーユニット、1 6…冷却ファン、2 0…チャネルアダプタ、2 1…通信ポート、3 0…ディスクアダプタ、4 0…キャッシュメモリ、5 0…共有メモリ、6 0…スイッチ部、7 0…論理デバイス、8 0…ディスクドライブ、9 0…サービスプロセッサ、9 1…認証部、9 2…構成管理部、1 0 0…キュー、1 0 1…フリーキュー、1 0 2…クリーンキュー、1 0 3…ダーティキュー、1 0 4…色分けキュー、1 0 5…BIND キュー、2 0 0…カウンタ、2 0 1…フリーキューカウンタ、2 0 2…クリーンキューカウンタ、2 0 3…ダーティキューカウンタ、2 0 4…色分けキューカウンタ、2 0 5…BIND キューカウ

ンタ、2 0 6…使用中カウンタ、3 0 0…キュー管理テーブル、CLPR…論理的キャッシュ
分割領域、SLPR…仮想筐体、SLCB…スロット制御テーブル、SGCB…セグメント制御ブロッ
ク、F 1…DCR、F 2…サイドファイル、F 3…PCR、CN…通信ネットワーク、S
T…ステータス、T 1～T 7…テーブル

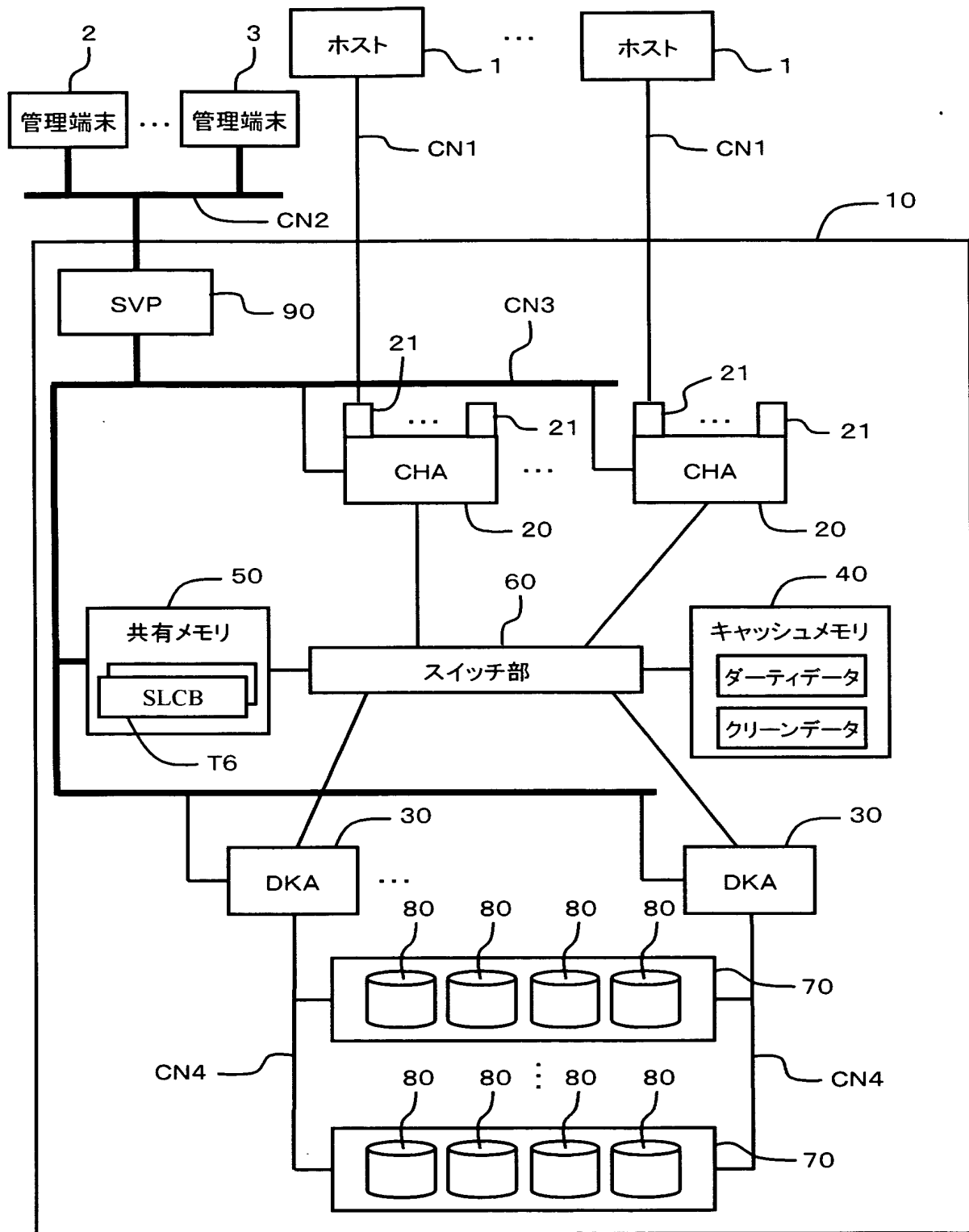
【書類名】 図面
【図 1】



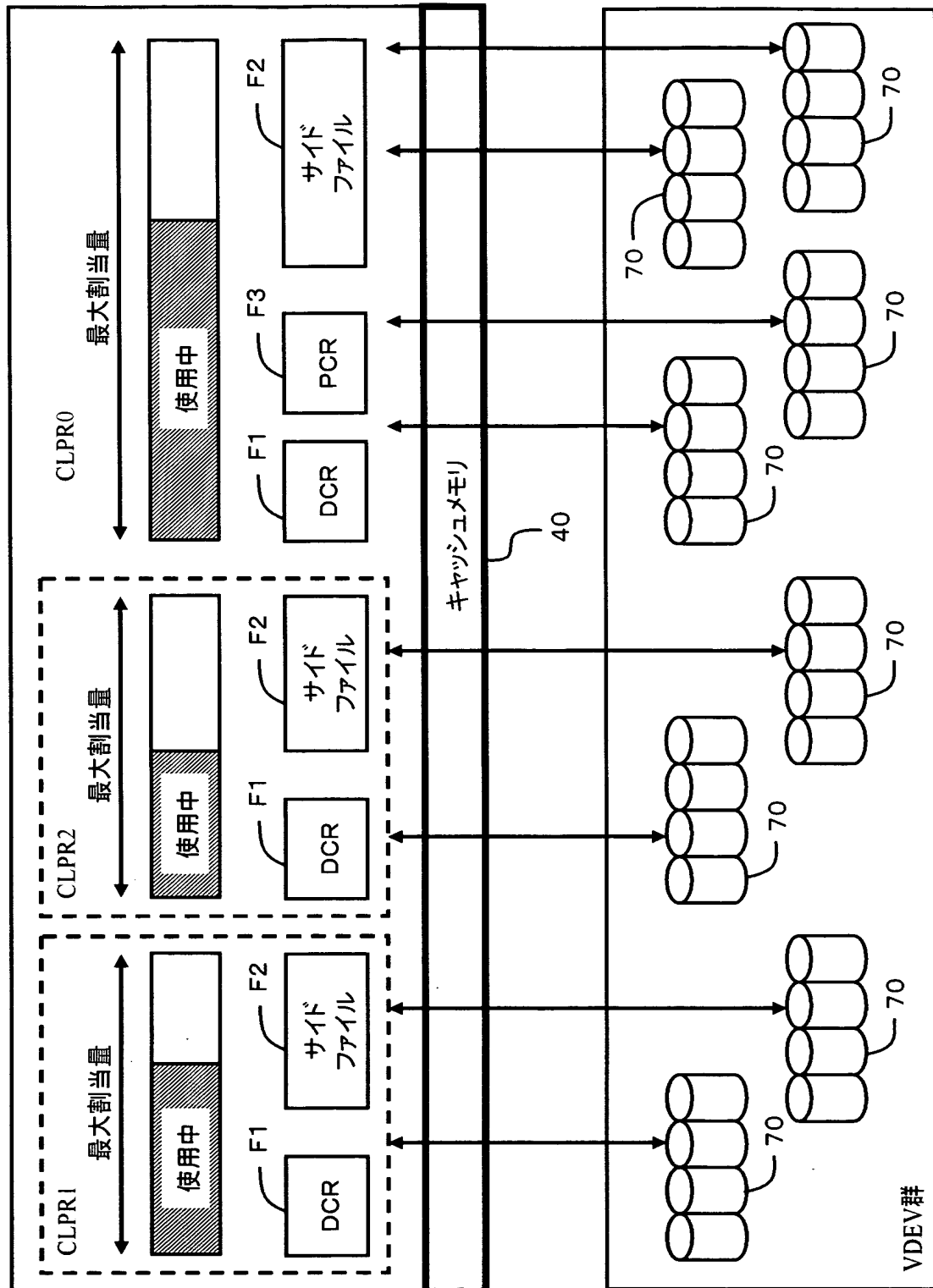
【図 2】



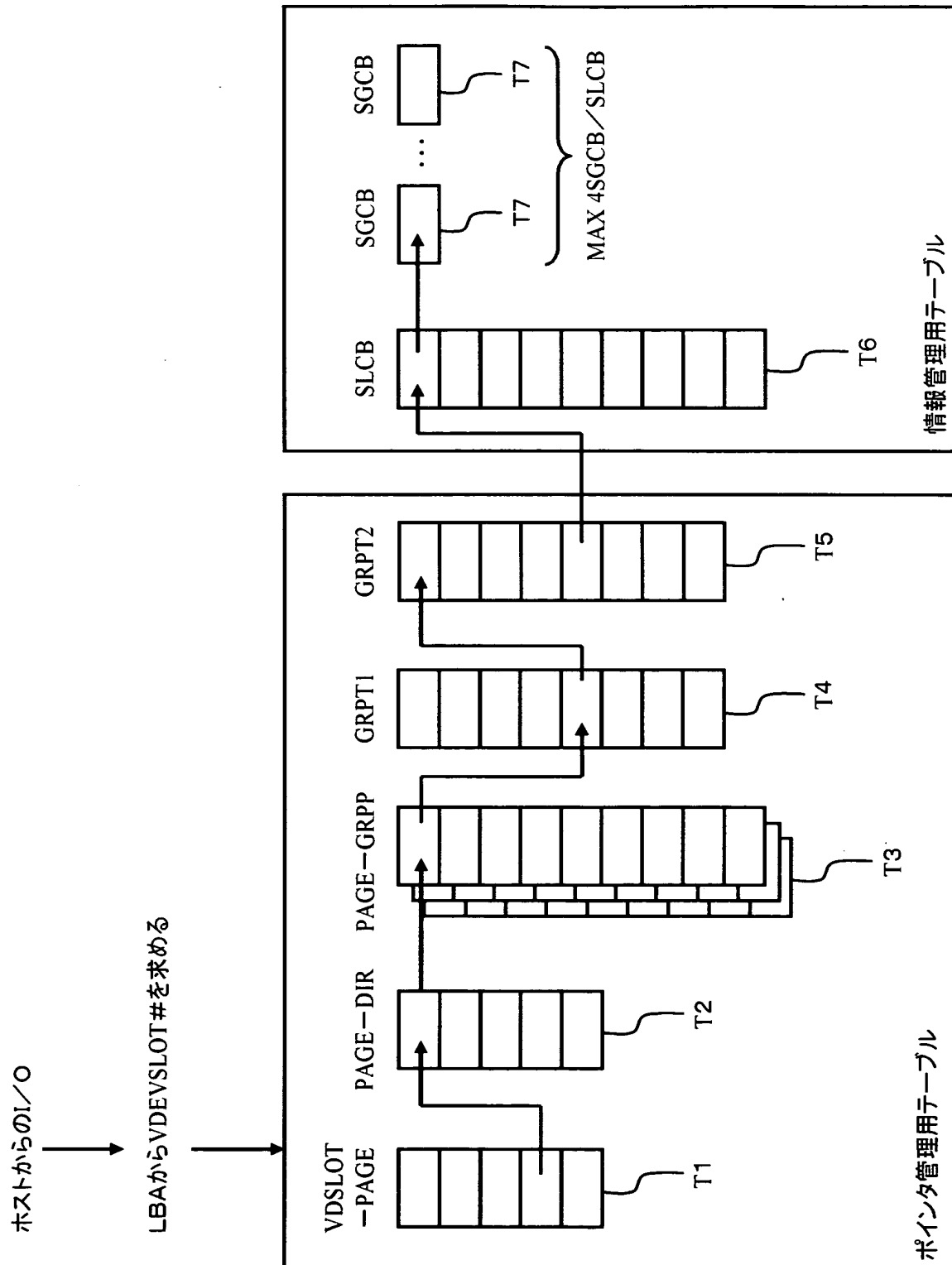
【図 3】



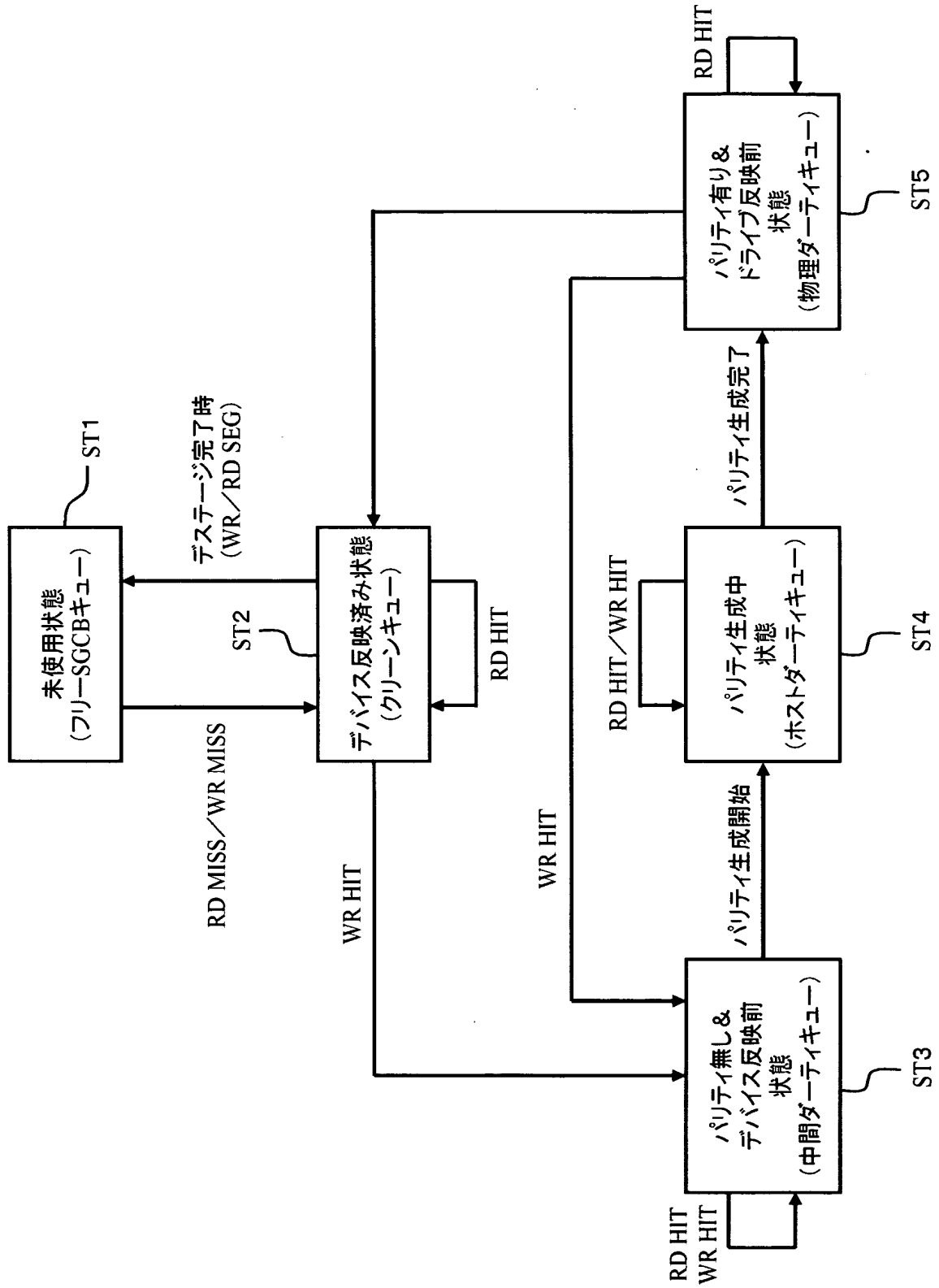
【図 4】



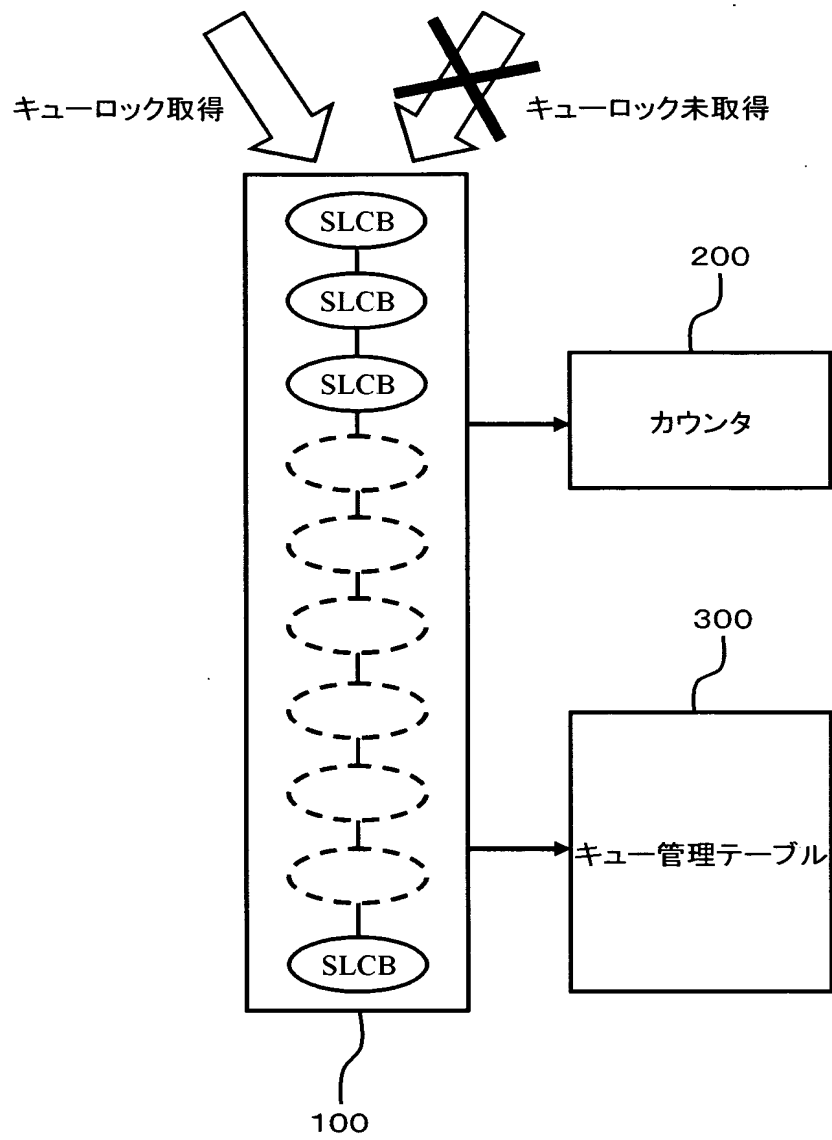
【図 5】



【図 6】



【図 7】

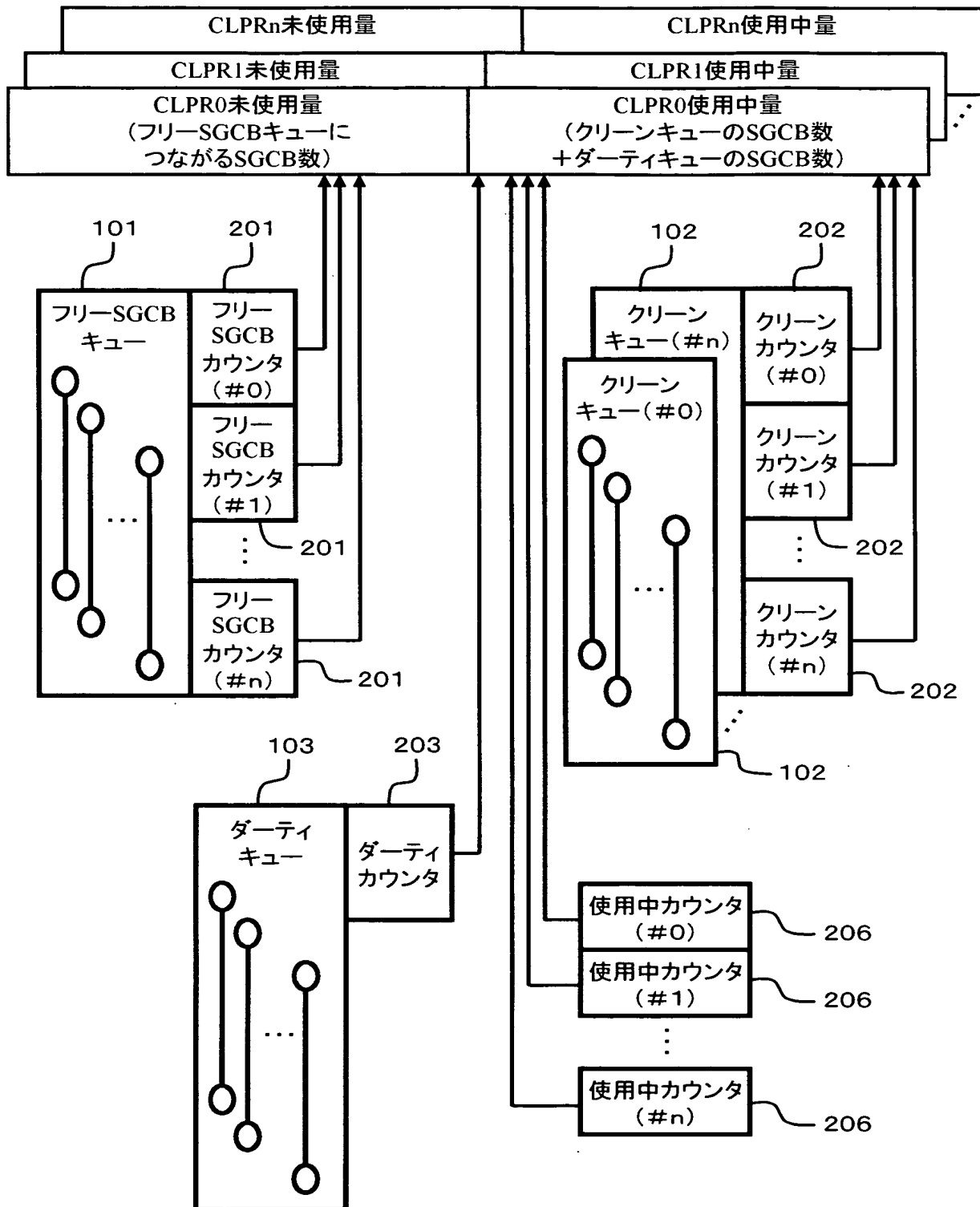


【図 8】

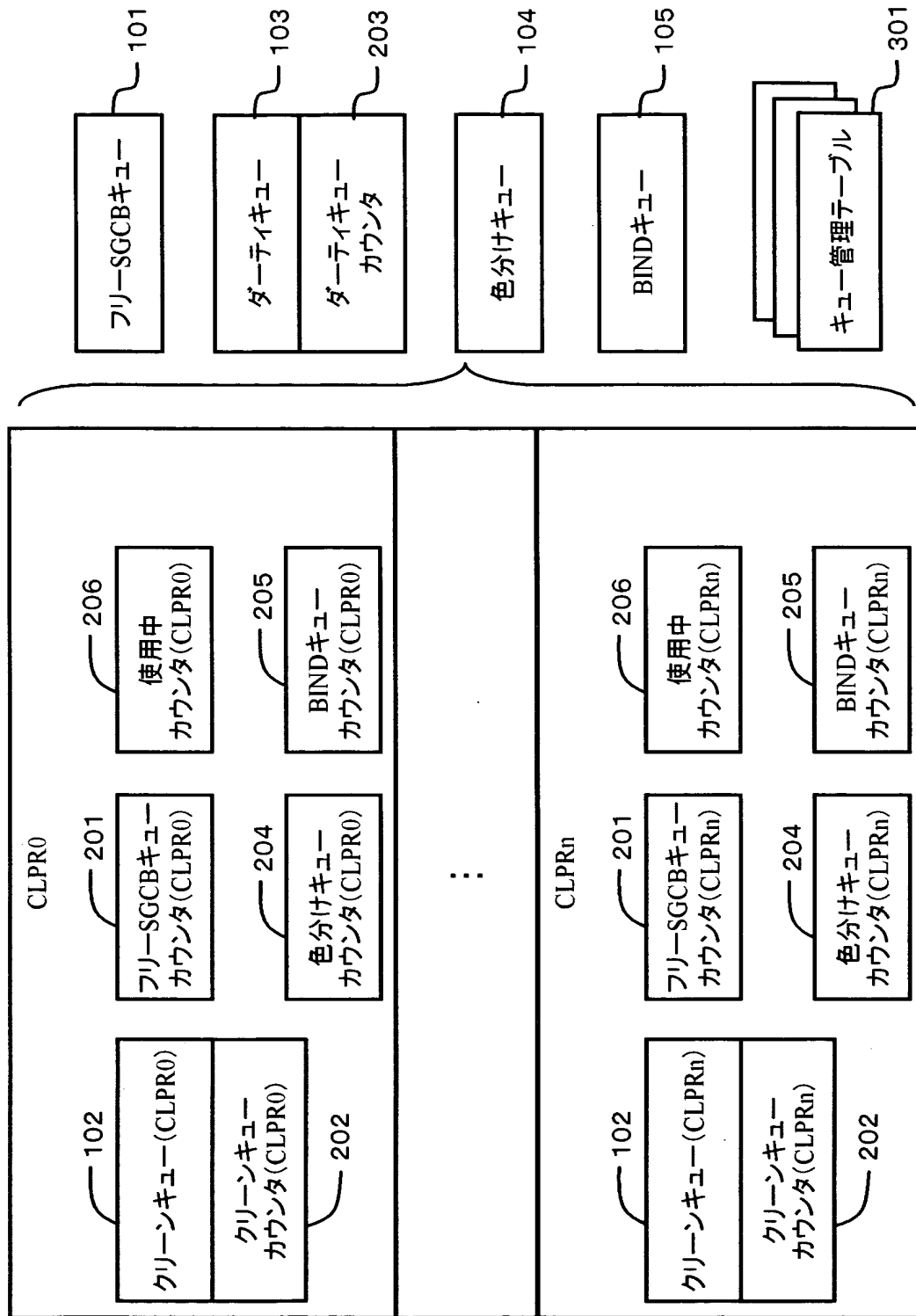
— T6

SLCBテーブル			
後方ポインタ	前方ポインタ	キューステータス (キュー種別、キュー#)	VDEVSLOT#
スロットステータス (SLCBの状態)	CLPR情報	SGCBポインタ	SGCBポインタ
⋮	⋮	⋮	⋮

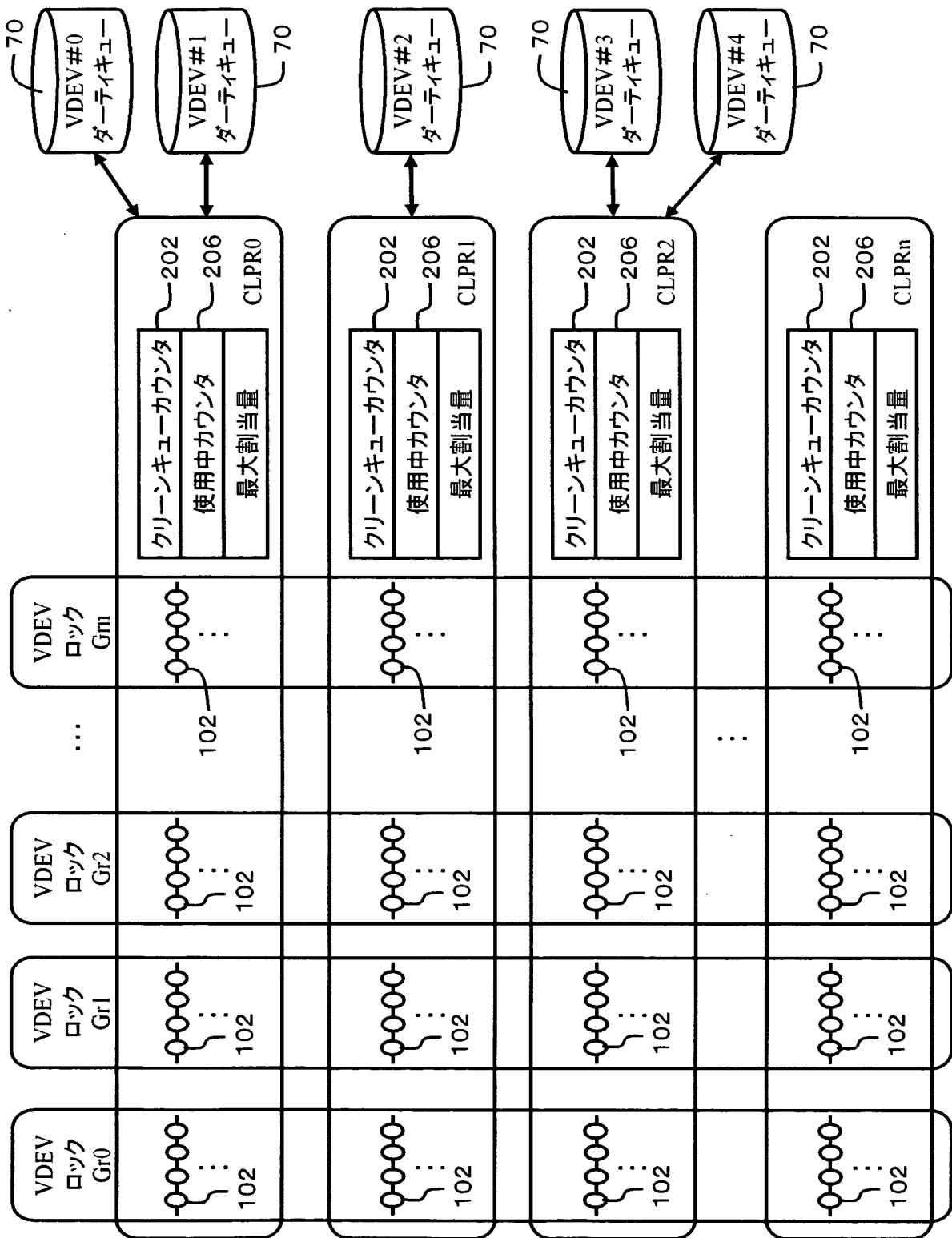
【図 9】



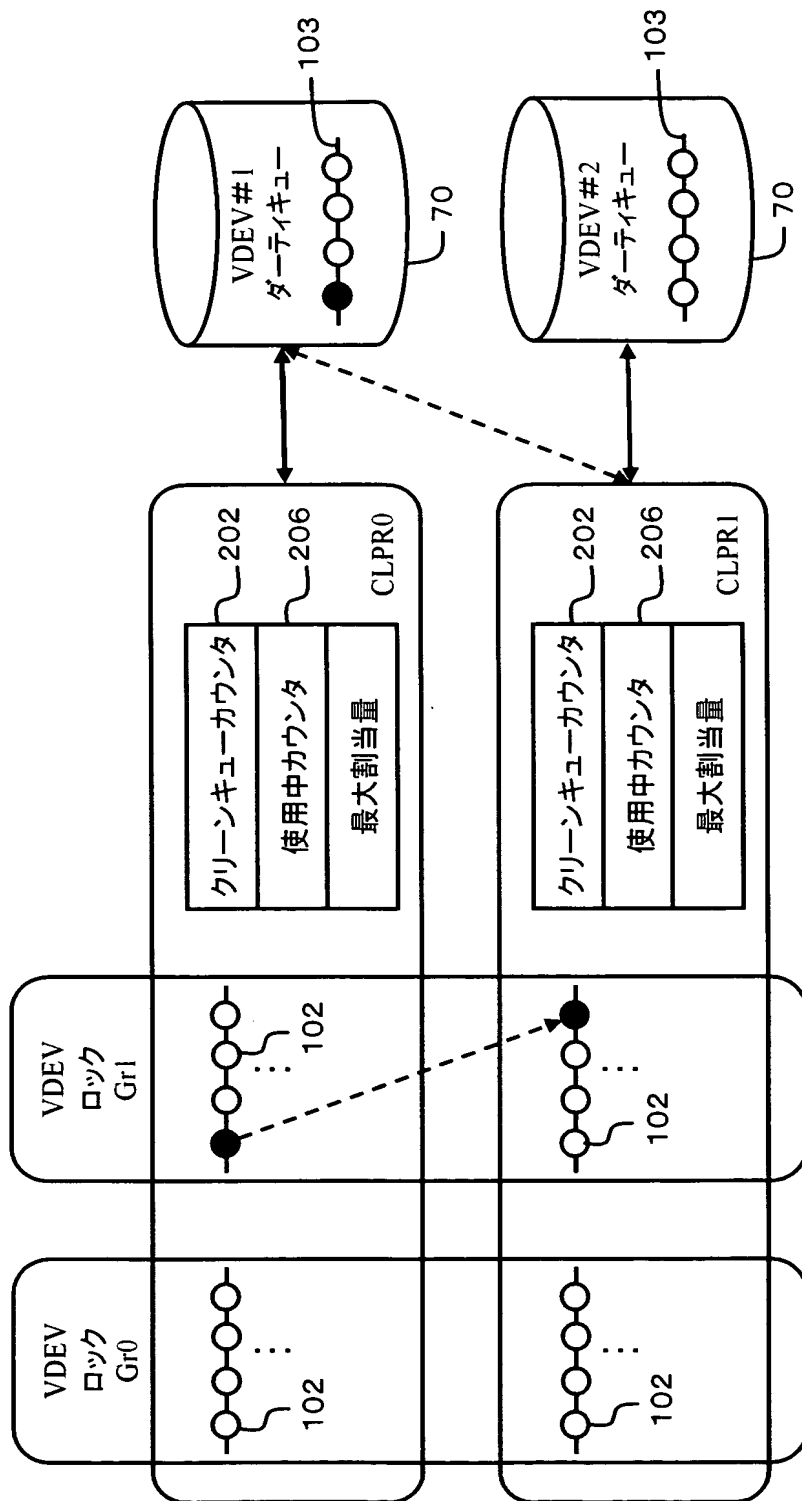
【図 10】



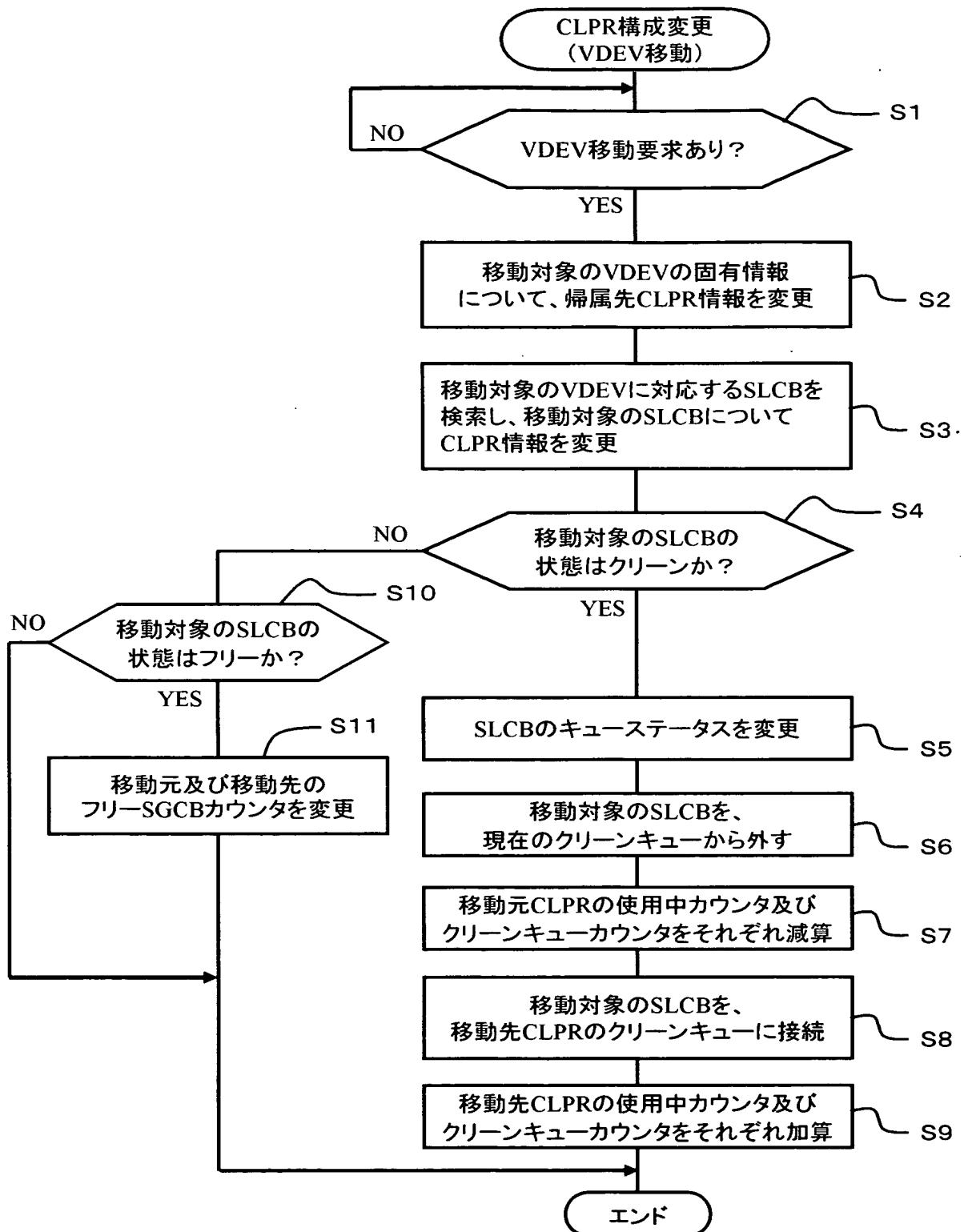
【図 11】



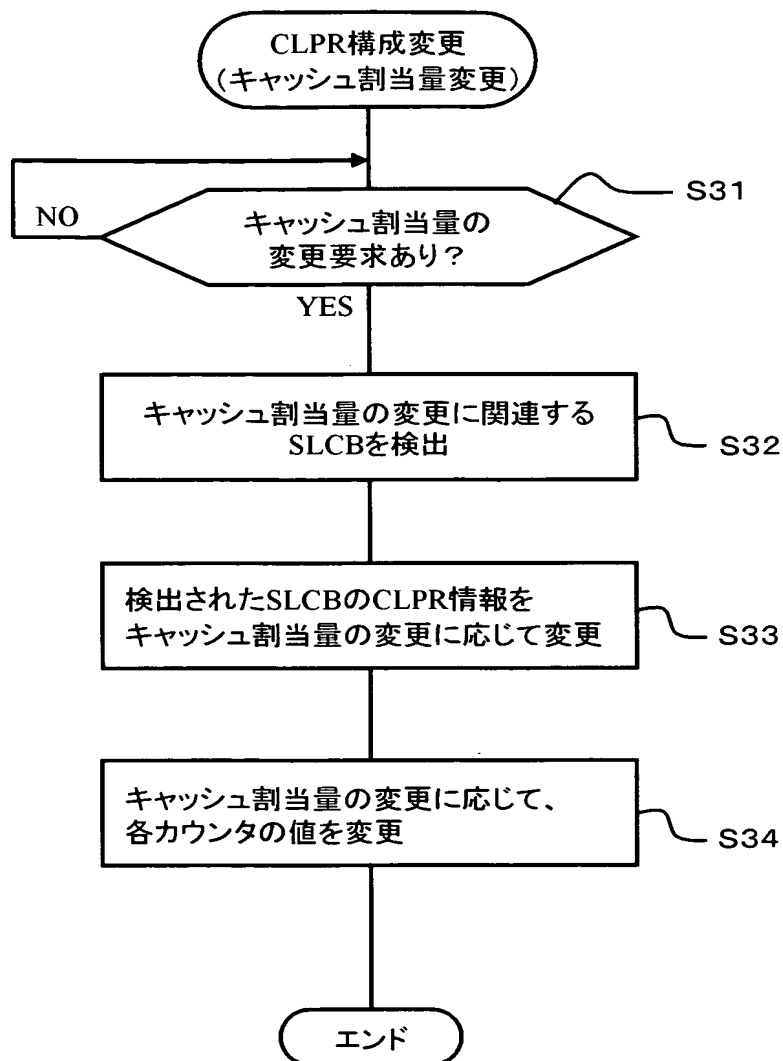
【図 12】



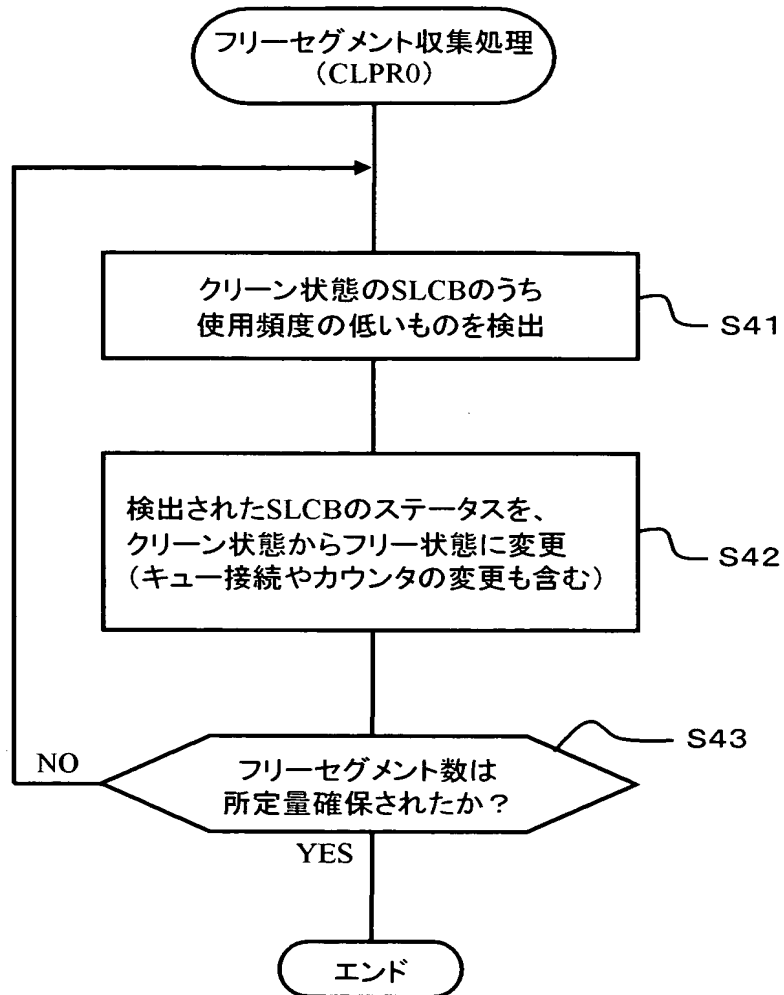
【図 13】



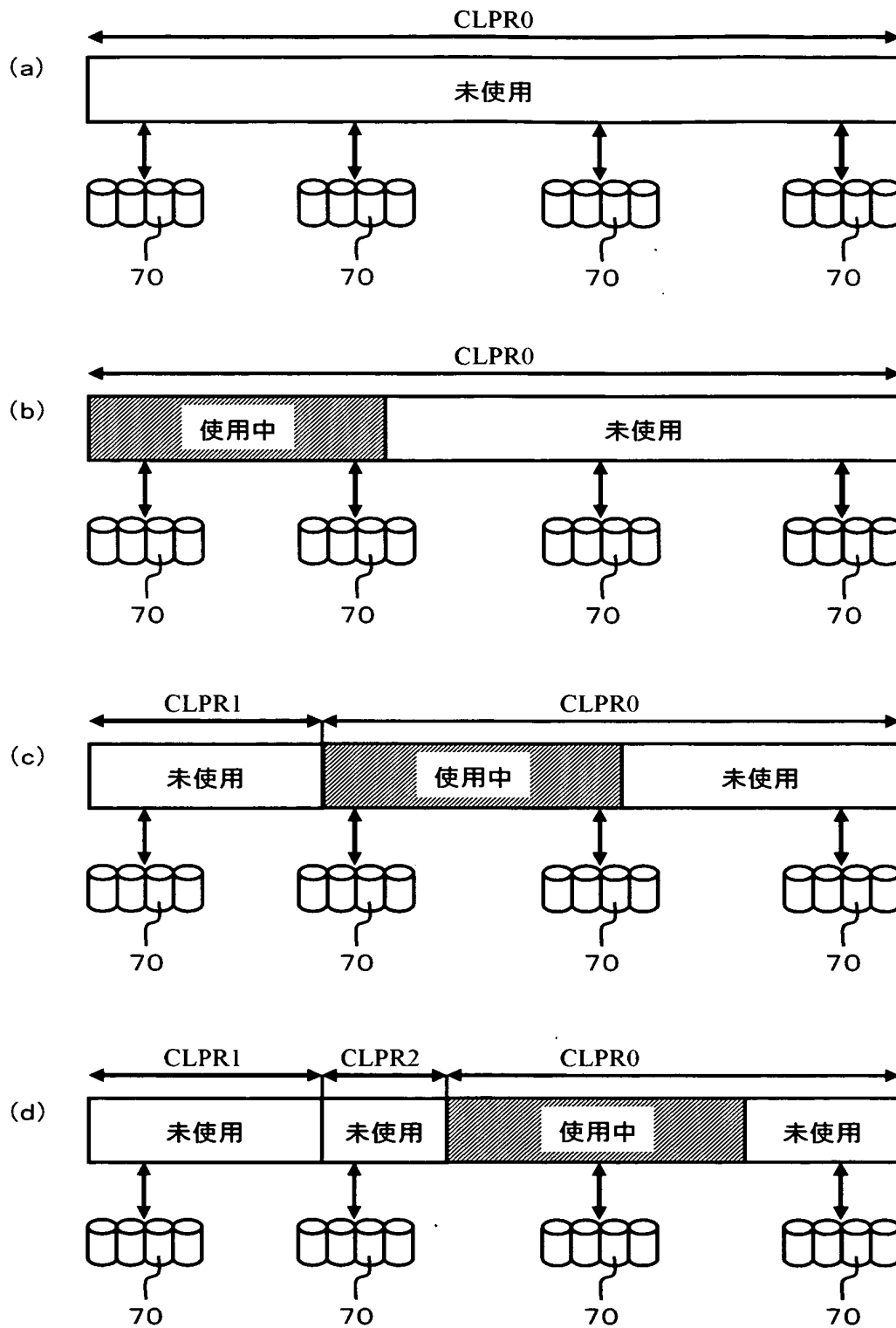
【図 14】



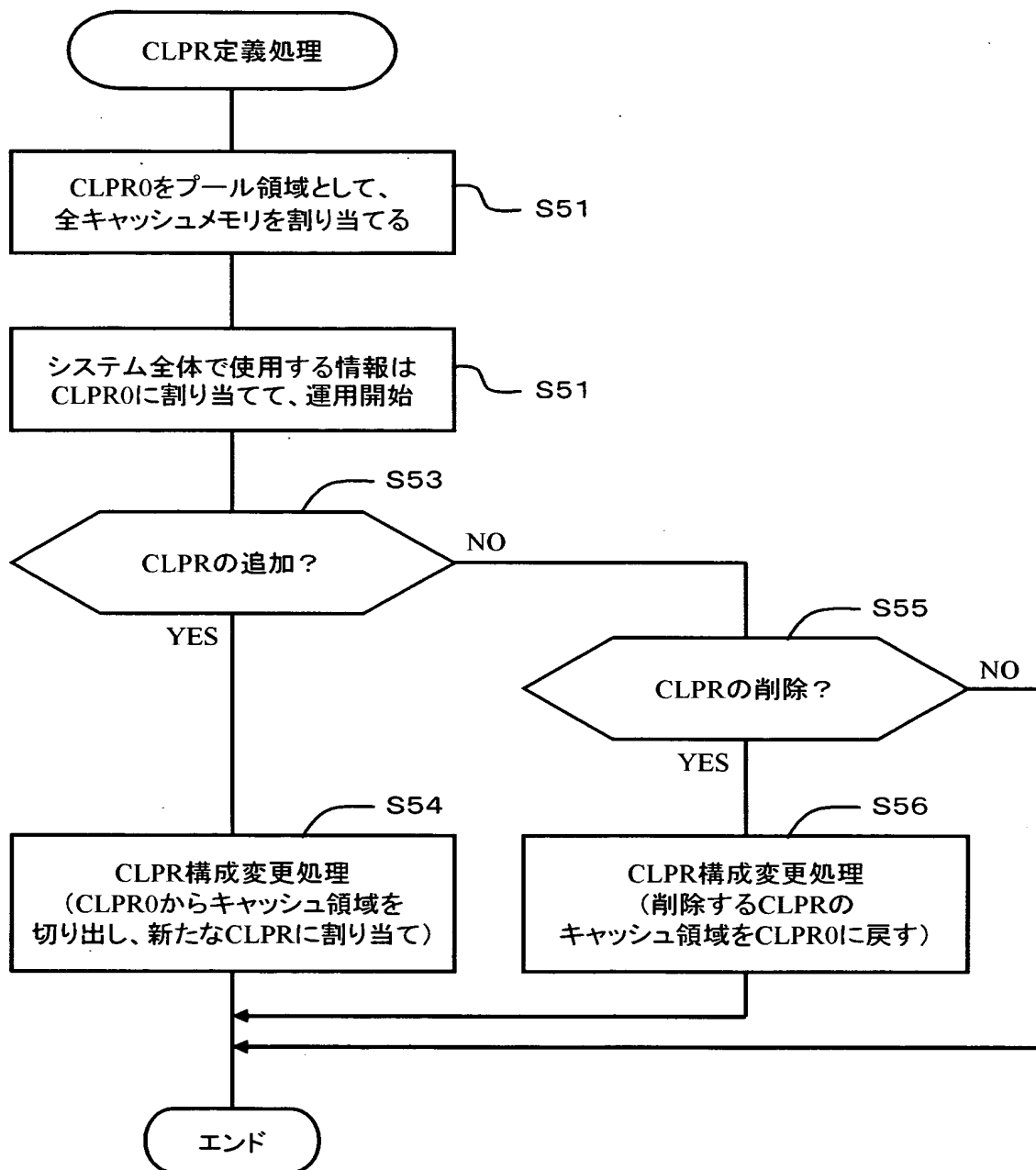
【図 15】



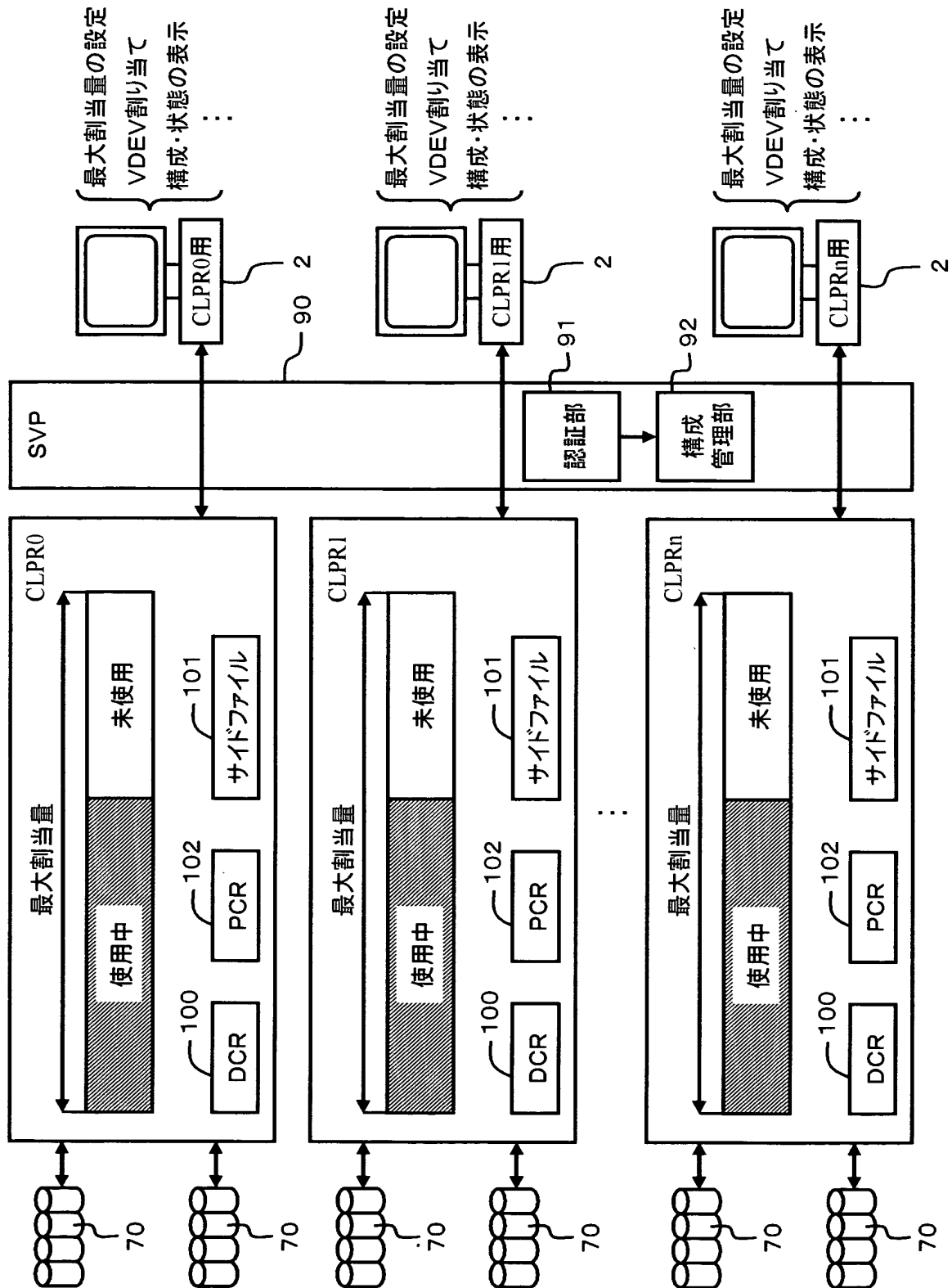
【図 16】



【図 17】



【図 18】



【書類名】 要約書**【要約】**

【課題】 ストレージサブシステムのキャッシュ領域を各ユーザ毎に分割し、ユーザ専用領域同士の干渉を防止する。

【解決手段】 ストレージサブシステム内には、複数のCLPRを設定可能である。CLPRは、キャッシュメモリの有するキャッシュ領域を分割して使用可能とした、ユーザ専用領域である。キャッシュメモリに記憶されるデータを管理するために必要な管理情報は、セグメントまたはスロットの属性に応じて、各CLPRに分配される。クリーン状態のセグメントを管理するクリーンキュー 1 0 2 及びクリーンカウンタ 2 0 2 は、各CLPR毎にそれぞれ設けられる。ダーティキュー 1 0 3 及びダーティカウンタ 2 0 3 は、全CLPRで共通に使用する。フリーキュー 1 0 1、色分けキュー 1 0 4 及びBINDキュー 1 0 5 は、全てのCLPRに共通に適用し、各カウンタ 2 0 1, 2 0 4, 2 0 5 のみを各CLPRにそれぞれ設ける。

【選択図】 図 1 0

認定・付加情報

特許出願の番号	特願 2 0 0 4 - 0 2 9 0 2 8
受付番号	5 0 4 0 0 1 8 8 4 5 8
書類名	特許願
担当官	第七担当上席 0 0 9 6
作成日	平成 1 6 年 2 月 6 日

< 認定情報・付加情報 >

【提出日】 平成16年 2月 5日

特願 2004-029028

出願人履歴情報

識別番号 [000005108]

1. 変更年月日 1990年 8月31日

[変更理由] 新規登録

住 所 東京都千代田区神田駿河台4丁目6番地

氏 名 株式会社日立製作所